CONTRACT NO. 94-333
FINAL REPORT
MAY 1998

# Temporal, Spatial, and Ambient Temperature Effects in the Sacramento Modeling Region

CALIFORNIA ENVIRONMENTAL PROTECTION AGENCY

AIR RESOURCES BOARD
Research Division

# TEMPORAL, SPATIAL, AND AMBIENT TEMPERATURE EFFECTS IN THE SACRAMENTO MODELING REGION

Final Report
Contract No. 94-333

Prepared for:

California Air Resources Board
Research Division
2020 L Street
Sacramento, CA 95814

Prepared by:

David M. Rocke, Director
Center for Statistics in Science and Technology

Daniel P. Y. Chang
Department of Civil and Environmental Engineering

University of California, Davis
Davis, CA 95616

**May 1998**

For more information about the ARB's Research Division,
its research and activities, please visit our Web site:

http://www.arb.ca.gov/rd/rd.htm

## Disclaimer

**Acknowledgments**

The University of California at Davis is pleased to acknowledge the participants of this study. Sonoma Technology Inc. (STI) provided invaluable assistance with the survey design and questionnaire development. Especially, Lyle Chinkin and Hilary Main at STI contributed their knowledge and experience in emission inventories to this project. We also gratefully acknowledge the assistance of the ARB staff including Robert Grant, Dale Shimp, Cheryl Taylor, Bob Weller, Patrick Gaffney, and Patricia Velasco. The California Department of Water Resources provided the land use data, and the University of California Cooperation Extension provided the data on sample production cost by major crops. The Visualization Laboratory at the University of California, Davis, provided assistance with GIS software and consultation.

This report was submitted in fulfillment of contract agreement 94-333, "Temporal, Spatial, and Ambient Temperature Effects in the Sacramento Modeling Region," by the University of California at Davis, under the sponsorship of the California Air Resources Board. Work was completed as of 10/31/97.

**Abstract**

The objective of this study is to identify and investigate important temporal and spatial variations in factors that affect emissions in the Sacramento modeling region, especially from non-road mobile sources and industrial surface coatings and related process solvents. In addition the project assesses the effect of ambient temperature on emissions from those sources. To accomplish the objective, this study develops spatial activity indicators and temporal activity profiles for the emission source categories included in the study. Specifically, a geographical information system (GIS) based approach is developed to spatially allocate regional or county-level emissions to units such as grid cells used in photochemical air quality simulation models. Statistical models are developed by which the values of spatial surrogates can be estimated and updated using widely available data such as those of land uses, population census, and the U.S. Census Bureau's topographically integrated geographic encoding and referencing (TIGER) files. Most temporal activity profiles developed in this study are based on our surveys of emission source facilities in the study area. The temporal activity profiles can be used to scale the annual emission estimated by the ARB to determine monthly, weekly (day of the week), and hourly emissions. The monthly activity profiles of the farm equipment category are calculated from crop-specific sample production cost estimates developed by the county farm advisors and the University of California Cooperative Extension. The effects of ambient temperature and weather on the source activities are estimated based on the data from our surveys.

**Table of Contents**

## List of Figures



## List of Tables

# Executive Summary

Area source emission inventories require estimating the amount of emissions for various industrial, commercial, and consumer activities at different spatial and temporal scales. The basic methodology was developed in the 1970s as a part of the National Emissions Data System, by which national solvent usage is allocated to the States and counties. The lack of emission estimates at subcounty levels and various temporal scales severely limits the modeling and planning capabilities in urban and regional air quality management. This study develops new methods of spatial allocation, temporal activity profiles, and estimates of ambient temperature effects for selected emission inventory categories in non-road mobile sources and industrial surface coatings and related process solvents. Specifically, six source categories are included in this project: auto refinishing, adhesives and sealants, can & coil coatings/metal parts and products coatings, farm equipment, construction equipment, and trains. The study region consists of three counties in California: Sacramento County, Solano County, and Yolo County.

In this study temporal activity profiles and ambient temperature effects are estimated largely based on data from our surveys of industrial and commercial facilities. A new method is developed to spatially allocate countywide emission estimates to model grid cells. The spatial allocation procedure consists of the following steps: for each source category, selecting an activity indicator, estimating the level of activity in each cell, computing allocation factors, and allocating emission estimates to the cells. A geographical information system (GIS) is used to create and process the spatial data and to carry out the spatial allocation. To reduce the effort on data collection, statistical models are used to correlate activity indicators with easily available data and to predict the level of activities. The unit of spatial allocation in this study is a 4 km by 4 km grid cell. As requested by the ARB, the allocation surrogates chosen in the study should be based on data that can be easily collected and regularly updated. Thus, only simple activity indicators are used in the allocation, such as the number of emission producing facilities per cell, or miles of railroads per cell. The variables selected for predicting the level of activities are also very simple, based on data widely available such as population density and employment statistics (census data), land uses (available from state or local governments), and major roads and highways (the TIGER files).

For spatial allocation, this report provides activity indictors, the procedure of allocation, and the regression equations that can be used to estimate spatial distributions of activities based on widely available data. For temporal allocation, monthly, weekly (day of the week), and diurnal activity profiles are provided. Since emissions are estimated by multiplying emission factors with activity levels, the temporal activity profiles developed in this study can be used to scale annual emission estimates to determine monthly, weekly, or hourly emissions. Specifically, the weekly profiles contain the fraction of

weekly emissions allocated to each day of the week; the hourly profiles provide the fraction of daily emissions allocated to each hour of the day. The confidence intervals for the estimates are presented. The impact of temperature and weather on the source activities are also assessed. Two effects are investigated: (1) direct effect of increasing temperature or raining; and (2) indirect effect of changes in activity patterns which demonstrate significant time shifts to account for high ambient temperatures. Percentage of changes in activities due to those effects are presented and confidence intervals of the estimates are provided.

Information on spatial and temporal distributions of emissions is essential for developing emission inventories and ozone air quality simulation models such as the Sacramento State Implementation Plan (SIP) Urban Airshed Model (UAM). This study contributes to improvement in area source emission inventories by developing new methods of spatial allocation, allocation surrogates, temporal activity profiles, and estimates of ambient temperature effects. Recommendations for further study include development of better allocation surrogates, collection of additional data, and improvement of accuracy of the estimates.

# 1. Introduction

The purpose of this project is to identify and investigate important temporal and spatial variations in factors that affect emissions in the Sacramento modeling region, specifically from non-road mobile sources and industrial surface coatings and related process solvents. In addition, this project assesses the effect of ambient temperature on emissions from those sources. Information on temporal and spatial distributions of emissions is important to air quality monitoring, emission inventory development, and air quality simulation and modeling. For example, photochemical air quality simulation models such as the Urban Airshed Model (UAM) used in the Sacramento State Implementation Plan (SIP) applications require detailed data on spatial and temporal distributions of emissions in the study area.

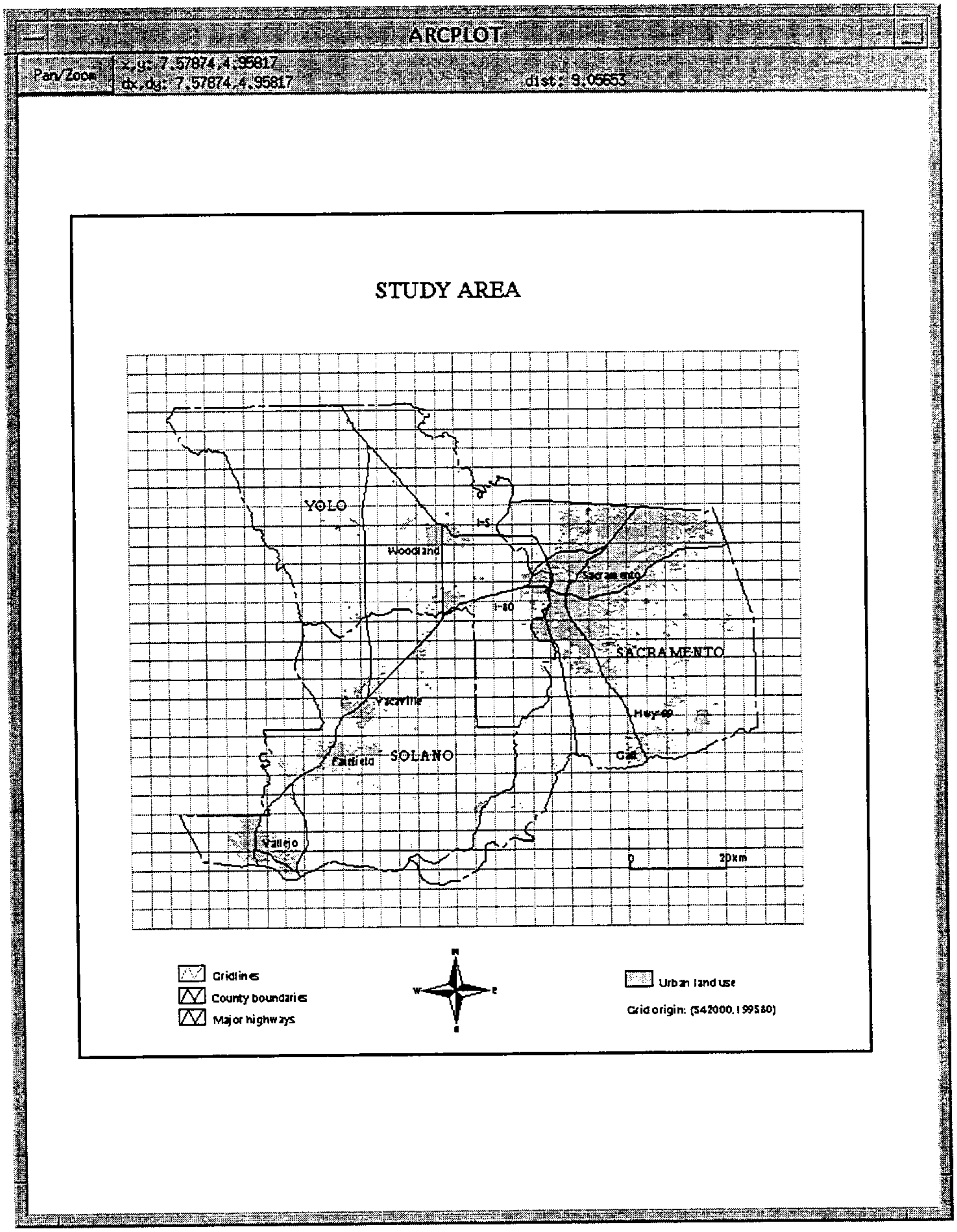
The following emissions source categories were selected to be included in this project:

- Auto Refinishing
- Adhesives and Sealants
- Can and Coil Coatings / Metal Parts and Products Coatings
- Farm Equipment
- Construction Mobile Equipment
- Trains

The study area includes three counties in the State of California: Sacramento county, Solano county, and Yolo county. The study area is shown in Figure 1.1.

In selecting emission source categories for use in this study, the California Air Resources Board (ARB) considered several factors: (1) the amount of ozone precursor emissions from the categories relative to total emissions in the Sacramento modeling region and statewide, and (2) the lack of readily available data. Table 1-1 provides a summary of the current approaches by U.S. EPA and/or the ARB to estimate activity indicators and spatial distributions for the source categories selected. As shown in Table 1-1, activity indicators for a number of categories are based on industrial surveys, equipment sales, fuel usage and from the 1982 Census of Manufacturers. Table 1-1 also shows that spatial allocations are generally based on a single indicator such as land use types (e.g., urban or rural) or some population statistics. As part of this study we worked with the ARB to identify those categories for which the existing sources of activity and spatial indicators need to be updated with new surveys or data collection efforts and categories for which new indictors should be developed.

*Figure 1.1 Map of the Study Area*



This project has five tasks: development of temporal activity profiles, identification and assignment of spatial surrogates, identification and prioritization of categories affected by ambient temperature, development of estimates of the effects of temperature on emissions, and estimation of confidence intervals. These five tasks were performed along the following three lines.

**Development of Spatial Allocation Surrogates**

Area source emissions are usually estimated at the national or state level and then allocated to counties (CARB, 1995). Local variations of emissions within a county are rarely known. Consequently, emission inventories used in photochemical models such as the UAM have been spatially allocated to model grid cells using only readily available spatial allocation surrogates, such as population, housing, land use and some employment statistics available by census tract or land use zone (e.g., Rao, 1987; Scheffe, 1990; Morris and Myers, 1990). An alternative approach is to estimate areawide emissions using data collected at the individual level (e.g., Causley, 1995; Shimp and Campbell, 1996). While the bottom-up approach has great potential to provide more accurate emission inventories, the cost of data collection and the availability of data are the limitations.

In this project we developed a new approach to spatial allocation of area source emissions to sub-county units, specifically, model grid cells. This approach consists of selecting a spatial activity indicator for a given source category, examining the activity levels and spatial distribution, and estimating factors for spatially allocating county-wide emissions estimates to model grid cells. In the approach the spatial activity indicator selected for a source category is the spatial surrogate for allocation. As required by the ARB, surrogates selected in this study must be based on parameters that are collected and updated on a regular basis and that do not require the development of specialized data. We were also advised to select surrogates or parameters that are readily available for all areas of the Sacramento modeling domain (and statewide) to prevent discontinuity in emissions due to abrupt changes in surrogates that are simply artifacts resulting from changing surrogates at a county boundary. In this study we identified methods that can be used to estimate spatial surrogates based on widely available data.

**Development of Temporal Activity Profiles**

As noted above, the ARB selected categories for which there was a marked lack of readily available sources of data. Table 1-2 provides a summary of temporal activity profiles currently used by the ARB for the source categories included in this study (CARB, 1995). In this study we developed monthly, weekly (day of week), and diurnal activity profiles for most of the categories included in the study. Since emissions are estimated by multiplying emission factors with throughput or activity level, the activity profiles developed in this study can be used to scale annual emission estimates to determine monthly, weekly, or hourly emissions.

Due to lack of available data for determining temporal activity patterns, we relied heavily on surveys to collect data. The survey efforts included identifying survey subjects, selecting samples, designing appropriate questionnaires, conducting telephone interviews, and analyzing the data. Using the data collected we computed the fractions for allocating annual emissions to three time-scales: monthly, weekly, and hourly. We also calculated the confidence intervals for the data.

**Development of Estimates of Ambient Temperature Effects**

The surveys included questions about the effects of temperature (90°F and above) and weather (raining or not) on the activities of interest. Using the survey data we estimated the temperature and weather effects, and computed confidence intervals for the estimates. Two effects of temperature and weather were considered: (1) the direct effect of increased temperature or raining on the activity (e.g., the auto refinisher may change paint formulation on hot or rainy days), and (2) indirect effect of changes in activity patterns which may demonstrate significant time shifts to account for high ambient temperature (e.g., construction activities may be shifted from normal working hours to early morning to avoid the heat of the day). We identified the categories affected by those effects.

In summary we developed data and methods to spatially allocate countywide emission estimates to model grid cells and temporally allocate annual emission estimates to three time-scales (monthly, weekly, and diurnal) for the source categories selected for this study. We also evaluated the effects of ambient temperature and weather on those categories. The major assumption is that allocation of emissions for a source category can be made by using source activity indicators or activity levels. In particular, it is assumed that the spatial allocations can be made by using the activity indicators that were selected in this study (Discussions about the activity indictors are provided in the following sections). There are more assumptions about the statistical models and data, which will be discussed later.

We would also like to acknowledge the difficulties we encountered in collecting data. Although we made great efforts to make use of available data and to collect new data within resource limitations of this study, our efforts were not always successful. Our aim is modest; this study is but a first step toward improving emission inventories in accounting for temporal, spatial, and ambient temperature emission effects in the Sacramento modeling region.

The rest of the report is organized as follows. In section 2 we discuss the methodologies used for this study and sources of available data. In Sections 3-6, the work for each source category is described and results are presented. In Section 7 recommendations are provided.

*Table 1.1 Current ARB and/or EPA Approaches to Estimating Activity Indicators and Spatial Surrogates*

| CATEGORY | CURRENT APPROACH TO ESTIMATE ACTIVITY INDICATORS | CURRENT APPROACH TO SPATIAL SURROGATES |
|---|---|---|
| Auto Refinishing | National production data from the 1982 Census of Manufacturers of Paint and Applied Products allocated to California by ratio of CA to US vehicle registration and survey of auto refinishing shops. | Land use - Urban |
| Adhesives and Sealants | National production data from the 1982 Census of Manufacturers of Paint and Allied Products allocated to California by ratio of CA to US population and distributed to counties based on new construction trend data. | Land Use - Urban (Industrial)<br>Population (Non-industrial) |
| Can and Coil Coatings | National production data from the 1982 Census of Manufacturers of Paint and Allied Products allocated to California by ratio of CA to US population and distributed to counties based on non-retail employment. | Land Use - Urban |
| Metal Parts and Products | National production data from the 1982 Census of Manufacturers of Paint and Allied Products allocated to California by ratio of CA to US population and distributed to counties based on non-retail employment. | Land Use - Urban |

*Table 1.1 Continued*

| CATEGORY | CURRENT APPROACH TO ESTIMATE ACTIVITY INDICATORS | CURRENT APPROACH TO SPATIAL SURROGATES |
|---|---|---|
| Mobile Equipment: Heavy-Duty Industrial Equipment | Population of industrial /logging/construction/mining equipment based on industrial surveys and records of engine sales. | Land Use - Urban for construction and industrial gasoline, rural for logging, mining, and industrial diesel |
| Mobile Equipment: Light-Duty Industrial Equipment | Population of light industrial equipment based on industrial survey and records of engine sales. | Land Use - Urban |
| Mobile Equipment: Heavy-Duty Farm Equipment | Population of farm equipment based on records of engine sales, and information from manufacturers and trade organizations. | Land Use - Agriculture |
| Mobile Equipment: Light-Duty Farm Equipment | Population of light farm equipment based on information from manufacturers and trade organizations. | Land use - Agriculture |
| Trains: Road-hauling | Trains miles traveled, tons of trailing cars, duty cycles used to determine horsepower requirements and fuel use. | Land Use - Urban |

*Table 1.2 Temporal Activity Profiles Currently Used by the ARB*

| CATEGORY | TEMPORAL ACTIVITY PROFILES |
|---|---|
| Auto Refinishing | The annual activity is uniform. The weekly activity occurs on the five weekday days. The daily activity occurs 8 hour per day. |
| Adhesives and Sealants | N/A |
| Can and Coil Coatings | N/A |
| Metal Parts and Products | N/A |
| Mobile Equipment: Heavy-Duty Industrial Equipment | The annual activity is uniform. The weekly activity occurs on the five weekday days. The daily activity occurs 24 hour per day, with the majority of the activity occurring during daylight hours. |
| Mobile Equipment: Light-Duty Industrial Equipment | Same as above |
| Mobile Equipment: Heavy-Duty Farm Equipment | The annual activity increases during the spring and fall and uniform for the rest of the year. The weekly activity is nearly uniform with slightly lower activity on weekends. The daily activity occurs during daylight hours. |
| Mobile Equipment: Light-Duty Farm Equipment | N/A |
| Train | N/A |

# 2. Materials and Methods

This section provides discussions on the methods used in this study. In addition it describes the general data used in this project.

## Overview

As stated in the introduction section, the main purpose is to develop approaches and data to spatially allocate countywide emission estimates to model grid cells and to temporally allocate annual emission estimates to three time-scales (monthly, weekly, and diurnal) for the source categories selected for this study. As required by the ARB, the spatial indicators used in this study must be based on parameters or data that are widely available and can be updated on a regular basis. Except for the farm equipment category where monthly allocation factors can be estimated using data on production costs by crops, the temporal activity profiles were developed based on survey data. Thus, survey techniques were used. For spatial data development and spatial allocation, we relied heavily on GIS methods and developed a GIS-based approach to spatial allocations. To make spatial activity indicator values easily estimated and updated, we applied statistical models by which an activity indictor is correlated with data that are widely available such as population census and land uses. Statistical methods were also used to estimate confidence intervals for the estimated model coefficients and data on temporal activity profiles and temperature effects.

## GIS Methods

### The GIS Approach

A good description of a GIS and its characteristics is provided by Bachman et al. (1996) as below:

"A geographic information system (GIS) is a spatial analysis tool that can be used to model the interrelationships of geographic entities. A GIS consists of a data base containing spatially referenced land-related data as well as procedures for systematically collecting, updating, processing, and distributing that data. The fundamental base of a GIS is a uniform referencing scheme which enables data within a system to be readily linked with other related data. A true GIS can be distinguished from other systems through its capacity to conduct spatial searches and overlays that actually generate new information."

Theories and applications of geographic information systems are discussed in Laurini and Thompson (1992), Goodchild et al. (1993), Fotheringham and Rogerson (1994), Birkin et al. (1996), Goodchild et al. (1996), Longley and Batty (1996), and Quattrochi and Goodchild (1997), among others. See Shimp and Campbell (1996) for using a GIS to evaluate $PM_{10}$ area source emissions.

In this study we developed a GIS-based approach to spatially allocating countywide emission estimates to model grid cells. The approach consists of three main steps. The first step is to develop a spatial database. This includes identifying emission sources in the study area, selecting activity indicator(s), geo-code activity locations, and developing other geo-referenced data related to the activity. The second step is to disaggregate large areas such as counties into smaller zones (e.g., model grids) and then aggregate source activity at the individual level into the zones. Data to be used in the statistical models are converted to the same zone-based unit (e.g., number of people in the zone). In the third step spatial allocation factors are computed. Those factors can be used to allocate countywide emission estimates to the zones. The allocation factors are estimated based on the activity indicator whose values can be either determined from observations or estimated by statistical models using widely available data. The statistical models will be introduced later in this section.

### Spatial Database Development

The basic spatial objects in a spatial database are points, lines and polygons (areas). In the spatial database for this project the object for emission allocation is an area entity represented by a grid square, which is usually used in photochemical air quality simulation models. Each industrial or commercial facility that produces emissions is represented by a point entity consisting of a pair of XY coordinates with several attributes. The location of a facility can be identified and geocoded by its address, which can be found from a variety of sources such as phone books, government records, commercial business lists, or online databases.

Geocoding is a mechanism for building a database relationship between addresses and spatial features. In geocoding, a GIS compares the address of a facility against the ones on a digital street map. When a match is found, a geographic coordinate pair is calculated for the address and a spatial point is created in the database. If the address isn't matched, a GIS would give diagnostic messages that explain why the address is not matched. The address can then be edited, and the address-matching process restarts. Specifically, the ARC/INFO (Environmental Systems Research Institute, Inc.) GIS provides the following address-matching capabilities: creating an address coverage (or converting TIGER/Line or other street files to an address coverage), building and maintaining INFO files containing a list of addresses to be matched, matching the list to the address coverage to create points, processing unmatched addresses, and maintaining address coverages.

To create the model grid coverage, the cell size must be chosen. In urban airshed model applications the cell size is normally in the range of 2 to 10 km (Morris and Myers, 1990). In this project we selected 4 km as the cell size, which provides a reasonable spatial resolution for emission allocation. The map extent was chosen to cover the whole study area. The coordinate system used was the Universal Transverse Mercator (UTM) system. The coordinates for the intersections of the grid lines were computed using a program written in the Arc Macro Language (AML). The model grid coverage was built into polygon topology.

### Spatial Overlay and Allocation

The spatial database consists of a variety of elements in a variety of units. For example, in the database, the automobile refinishing facilities are spatial points, while the unit for allocation is grid squares. It is a common problem in spatial analysis that the spatial units used by available data are not necessarily the ones required by the analysis or modeling. A solution to the problem is spatial overlay. Three types of spatial overlay operations have been used in this project:

- point in polygon operation (e.g., overlaying the point coverage of emission source locations with the polygon coverage of model grids);
- polygon on polygon operation (e.g., overlaying the census tract based polygon coverages with the polygon coverage of model grids);
- line in polygon operation (e.g., overlaying the line coverage of railroads with the polygon coverage of model grids).

All these operations have been done using the ARC/INFO GIS. Among the operations polygon overlay is more complicated than the other two. A technical issue involved in polygon overlays is known as the areal weighting problem. Since the boundaries of the source zone and the target zone usually don't coincide, one must weigh the source zone values according to the area of the target zone they make up. The method used to proportion a polygon's (source zone) attribute value to a model grid (target zone) is briefly described below.

### Area Weighting Method

Let V be the variable of interest, S be the source zone, T be the target zone, and A be the area of a zone. In the example of deriving population density for the model grid cells, V is the population density variable, Zone S may be a census tract, and Zone T is a grid cell. As S intersects T, their boundaries form a zone of intersection ST. The problem is finding the value of V for the target zone or the intersection zone. The computation depends on the measurement of V, whether it is "extensive" or "intensive," as suggested in the spatial analysis literature (e.g., Goodchild and Lam, 1980). The variable V is extensive if its value for a target zone is equal to the sum of its values for the intersection zone. V is intensive if its value for a target zone is weighted average of its value for the intersection zones. V is usually considered to be extensive if it is a count (e.g., number of people in a census district), and to be intensive when it is proportions, percentages or rates (e.g., percentage of urban land).

Assuming that V is evenly distributed within the source zone, the values of V are computed as follows. If it is an extensive variable, equation (2.1) applies

$$V_t = \sum_s \frac{V_s A_{st}}{A_s}. \tag{2.1}$$

If V is an intensive variable, equation (2.2) can be used

$$V_t = \sum_s \frac{V_s A_{st}}{A_t}. \tag{2.2}$$

The assumption that the variable of interest is uniformly distributed over a source zone is not always plausible. For example, there might be a lake in the zone. In those cases the methods of areal interpolation using ancillary data (Green, 1990; Flowerdew and Green, 1994) might be used, which takes into account other relevant information available about the source zones.

**Methods of Calculating Spatial Allocation Factors**

Spatial allocation factors can be calculated using the spatial activity indicator selected for a particular emission category. Two calculation methods can be used depending on the measurements of the indicator. The first method can be used when the indicator is considered to be discrete. In this method, the facility point coverage is overlaid with the model grid coverage, and then the allocation factor (fraction) for each grid cell is computed. Let $a_{ij}$ be the value of attribute of interest of point j (j=1,...,M) in cell i (i=1,...,N), and $w_i$ be the factor for allocating countywide emission estimates to cell i. The allocation factor for cell i is computed by

$$w_i = \frac{\sum_{j=1}^{M} a_{ij}}{\sum_{i=1}^{N}\sum_{j=1}^{M} a_{ij}}. \tag{2.3}$$

This method treats the modeling area as a collection of discrete grid cells. The method is simple and efficient but the results cannot be used to produce a smooth emission surface. If the purpose is to map the spatial distributions and the activity indicator is measured at the interval or ratio scale, an alternative method may be used. In the second method, the values of the indicator are used to interpolate a surface; the surface is then overlaid with the model grid coverage, and finally the allocation factor for each cell is computed. The surface of interest can be generated by using the Triangulated Irregular Network (TIN) data model (Peuker et al., 1978; ESRI, 1994). The TIN is a surface model that uses a sheet of continuous, connected triangular facets based on a Delaunay triangulation of irregularly spaced sample points. In this project the first method was used and the spatial allocation factors were computed using ARC/INFO and its AML.

## Statistical Methods

In this project we used regression models and Poisson regression models to estimate the spatial allocation surrogates (activity indicators) for several of the source categories. We also computed confidence intervals for estimates on temporal activity patterns and temperature and weather effects. The statistical methods and their assumptions are briefly discussed below.

### Regression Models

The regression model is a general linear model, whose general form is

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k \tag{2.4}$$

where the variable denoted by $y$ is the response (or dependent) variable, the variables denoted by $x_1, x_2, \ldots, x_k$ are the predictor variables, and the regression coefficients are denoted by $b_0$, $b_1$, $b_2, \ldots$, $b_k$. The model expresses a relationship between the response variable and the set of predictor variables.

The regression model assumes that the relationship between the response and predictor variables may be expressed as

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \ldots + \beta_kx_k + e \tag{2.5}$$

where $\beta_0$, $\beta_1$, $\beta_2, \ldots$, $\beta_k$ are regression parameters and $e$ is the error term. The error term is used to account for all the variations in the response $y$ that is not modeled by the linear function of the predictor variables $x_1, x_2, \ldots, x_k$. In regression models the error term is assumed to be independent of one another and have a normal distribution with zero mean and standard deviation $\sigma$. The regression parameters are estimated using data on the response and predictor variables. The resulting estimates of parameters are called the regression coefficients. Equation (2.4) can be used to predict the values of the response variable, given values of the predictor variables and the regression coefficients.

### Poisson Regression Model

Poisson regression models are widely used for analyzing and predicting count data (e.g., number of pollution sources in a given area). Like the regression model discussed above the Poisson regression model assumes a relationship between the response variable and the predictor variables. However, the relationship is not linear. Let y be the response variable. The model assumes that each y is drawn from a Poisson distribution with parameter λ, which is related to the predictors $x_1, x_2, \ldots, x_k$. The Poisson model is given by

$$\Pr(Y=y)=\frac{e^{-\lambda}\lambda^{y}}{y!}, \tag{2.6}$$

where y = 0, 1, 2, ... The most common formula for λ is

$$\ln\lambda=\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_k x_k=\beta' x. \tag{2.7}$$

It can be easily shown that

$$E[y|x]=Var[y|x]=\lambda=e^{\beta' x} \tag{2.8}$$

where β is a vector of coefficients and x is a vector of predictor variables. Thus, the Poisson model is a nonlinear regression. Denote b as the estimated value of β and $\hat{y}$ as the prediction, then

$$\hat{y}=e^{b'x}. \tag{2.9}$$

The coefficients of the Poisson regression model can be estimated using the maximum likelihood method. The log-likelihood function is

$$\ln L=\sum_i\left[-\lambda_i+y_i\beta' x_i-\ln y_i!\right]. \tag{2.10}$$

The likelihood equation and the Hessian are

$$\frac{\partial \ln L}{\partial\beta}=\sum_i(-\lambda_i+y_i)x_i=0. \tag{2.11}$$

$$\frac{\partial^2 \ln L}{\partial\beta\,\partial\beta'}=-\sum_i\lambda_i x_i x_i'. \tag{2.12}$$

Statistical software provides fast, reliable and efficient estimation of the regression model and the Poisson regression model (e.g., Stata, S-Plus).

**Confidence Intervals**

In data analysis we often use sample data to estimate population values. Since the estimates can vary from sample to sample, we need information on the likely range of errors. Thus we use the sample data to construct an interval, [lower(X), upper(X)], such that we can expect this interval to contain the true parameter with some desired level of confidence.

Let $X_1, X_2, \ldots, X_n$ be a random sample, where $n$ is the size of the sample. Let $\overline{X}$ be the sample mean, $S$ be the sample standard deviation, and $1-\alpha$ be the confidence level. The confidence interval for $\overline{X}$ is

$$\overline{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}} \qquad (2.13)$$

where $t$ is the $t$ distribution with $n$-1 degrees of freedom.

Estimating confidence intervals for proportions $p$ (relative frequencies or percentages) needs a different formula. Let $Y$ be the frequency of measurements, then $p = Y/n$. The approximate $100(1-\alpha)\%$ confidence interval for $p$ is given by

$$p \pm z_{\alpha/2} \sqrt{\frac{p*(1-p)}{n}} \qquad (2.14)$$

where $z$ is the standard normal distribution. There are many reference books that discuss the statistical methods used in this study. See, for example, Hogg and Tanis (1988), Greene (1990), and Johnson and Wichern (1992).

## Survey Methods

### Selection of a Survey Mechanism

Widely used survey mechanisms include:

- Personal interview surveys
- Mail surveys
- Telephone surveys

Selection among these alternatives depends on a number of factors, including the purpose of the survey, the definition of sampling frame, the sample size, the types of questions to be asked, the likelihood of obtaining accurate answers, the length of the survey, and time and budget constraints. Each mechanism has advantages and disadvantages in specific situations.

The survey mechanism selected for this study is the telephone survey. The telephone survey has characteristics in common with both the personal interview and the mail survey and, for many applications, the best of both. There is a number of advantages with telephone surveys (Frey, 1983). In this project the telephone survey is used mainly because (1) it helps identify survey subjects, (2) it helps get higher response rates and reduce response time, (3) its costs are relatively low. As will be discussed later, of several source categories included in the study we do not know exactly what the population is. The telephone survey provides a quick way for us to identify the survey subject and to decide whether or not it is the right subject to survey.

### Sampling Issues

Prior to selecting a sample, the population of interest needs to be defined. This may be all automobile refinishing shops or all farm equipment in the study area. A list (or sampling frame) is needed of all of the sampling units within this population. In some cases this is a simple task. If a sample of all students in a public school is to be drawn, a listing is likely to be available. In many cases, however, a population list is difficult, if not impossible, to obtain. For example, it is difficult to identify all facilities in the study area that does metal surface coatings. Two problems with sampling frames are specifically relevant to this study. First, a sampling frame may contain missing elements. The telephone directories, for instance, include only those businesses and households that choose to be listed. The business lists available from commercial databases may also be incomplete. Second, a sampling frame may contain elements which are not part of the population for which inference is to be drawn. For instance, a random sample of major users of adhesives and/or sealants in the study area would be difficult to draw from a list with no indication of the nature of the businesses.

Given the sampling frame survey respondents are usually selected in some random or pseudo-random manner. A number of sampling techniques are available. Methods of probability sampling include simple random sampling, systematic sampling, stratified sampling, cluster or multistage sampling, and two-phase sampling. Methods of non-probability sampling include purpose sampling, quota sampling, snowball sampling and others. The simple random sampling method is used most frequently in practice.

Sample size determination for accurate estimation is difficult without some preliminary survey information. As an example, suppose that monthly factors are being determined. These are numbers between 0 and 1 that sum to 1, and would probably vary within a range of about 0.05 to 0.20. If a typical response has a standard deviation of 0.05, and the goal is to achieve 95% confidence intervals of length ±.01, then a sample would be needed of size about 100. This could be larger or smaller depending on the actual standard deviation. The size of the necessary sample is a major determinant of the survey cost, and thus of the number of surveys that can be run within the given budget. In general the size of a random sample is given by:

$$n = \left( \frac{Z\,\sigma}{e} \right)^2 \qquad (2.15)$$

where: n = sample size

Z = 1.96, for 95% confidence that a result lies within a given confidence interval

Z = 2.58, for 99% confidence that a result lies within a given confidence interval

$\sigma$ = Standard deviation of the sample

e = the desired size of the confidence interval, expressed as a decimal number.

### Questionnaire Design

Development of a questionnaire is an important part in survey design. A number of factors should be considered in questionnaire design, including the purpose of the survey, types of data and types of questions (e.g., Fowler, 1984). In addition one must balance the needs of different participants in the survey process and try to anticipate the reactions of respondents to various questions and procedures. In this study the basic questionnaire was designed by STI (Sonoma Technology Inc.). Questions included quarterly, weekly (Sunday-Saturday), and diurnal (morning, afternoon, evening, nighttime) activity patterns, the effects of high temperature (90°F and above) and weather (rain) on the activity. The questionnaire was used in the survey of automobile refinishers. The questionnaire was later modified to be used in the surveys for other source categories. The questionnaires were tested in pre-survey telephone interviews.

### General Data

The data and collection process of each individual source category will be described in the section for that category. Here we describe the data in general use. As required by the ARB, the spatial allocation surrogates must be based on parameters or data that are widely available and can be updated on a regular basis. Our approach is to correlate the surrogate with data easily available and estimate a regression function. Using the function the surrogate for spatial allocation can be easily updated or predicted. We collected and used following general purpose data for the study area:

- Land use data
- Census Bureau's TIGER files
- 1990 Population Census

The land use data were obtained from the California Department of Water Resources. The data were geo-referenced with attributes on land use types. We converted the data to ARC/INFO format. The documentation of the data is included in the Appendix. The land use data for the three counties in the study area were based on recent yet different year surveys. The data for Sacramento, Solano, and Yolo counties were from 1993, 1994, or 1989 surveys, respectively. We assume that land use data exist for almost all counties in the state. We used the TIGER files to obtain information on streets, highways, and railroads. The TIGER files are available from the U.S. Census Bureau and can be downloaded from the World Wide Web (http://www.census/gov/geo/www/tiger/). We used the 1994 TIGER files, but the files are updated on a regular basis and are available for the whole state and most of the country. We got the 1990 Population Census data from two sources. The census boundary (census tracts) files in ARC/INFO format were purchased from the Teale Data Center in Sacramento, which has the boundary files for all counties in the state. The census tract boundary files came with attributes on some population characteristics such as population density. The employment data were from the 1990 census CDs (Summary Tape File 3) and are available from many public libraries, for example, the UCD main library.

# 3. Auto Refinishing

## Source Activity and Sample Selection

The source category auto refinishing is used by the ARB to inventory the total organic gas (TOG) emissions that are from auto refinishing operations in California. The sampling frame we selected for this category is all auto refinishing shops in the study area. A list of auto refinishers in the three county areas was gathered from local phone books. The information on the list includes name of the shop, the address and phone number. All 344 shops on the list were used in the spatial analysis. The spatial distribution of auto refinishing shops in the study area is shown in Figure 3.1. A random sample of 78 shops was selected from the list for the telephone survey. The purpose of the survey was to develop temporal activity profiles and to assess the effects of ambient temperature.

## Spatial Surrogate for Allocation

### The Spatial Surrogate and Data

The spatial surrogate (spatial activity indicator) selected for this category is the number of auto refinishing shops in a given area, in particular, a 4 km by 4km model grid cell. Using equation (2.3) presented in Section 2 the spatial allocation factors can be directly computed. In this case all $a_{ij}$ in equation (2.3) are equal to one. Other spatial allocation surrogates can be used if data are available. For example, the spatial surrogate could be the number of employees in the auto refinishing shops or the number of cars painted by the shops within the area defined by the grid cell. The method for spatial allocation is the same. The selection of a spatial surrogate, however, depends on the availability of data.

The spatial surrogate can be estimated using data that are widely available and can be updated on a regular basis. Statistical models can be used for this purpose. The type of model to be used depends on the measurement of the spatial surrogate. Since the spatial surrogate is a count variable in this case, the Poisson regression model is used. In the model the dependent variable (AUTOREF) is the number of auto refinishing shops in a grid cell. The predictors are percentage of urban land use (URBAN), miles of highways (HWY), population density (POP), and retail employment density (RETEMP). Table 3.1 lists the variables used in the Poisson regression models. The sources of data for the predictors have been described in Section 2. To obtain grid based data, all spatial data layers (location of auto refinishing shops, land use, highways, population census) were overlaid with the model grid layer. Using an AML program we wrote, grid based data values were computed. The data were exported from the GIS database and then were read into a statistical program. Poisson regression models were estimated.

*Figure 3.1 Spatial Distribution of Auto Refinishing Facilities*

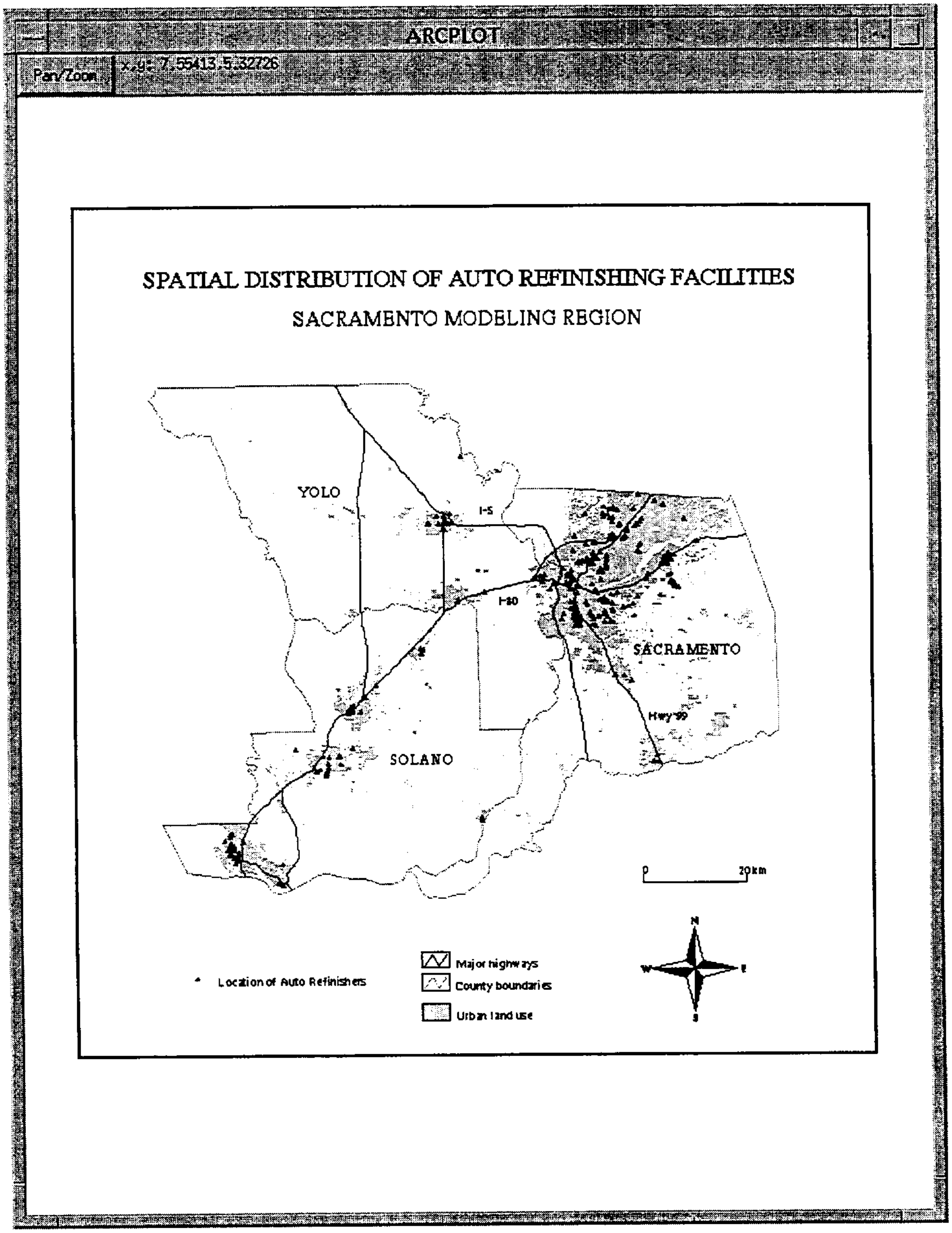


*Table 3.1 Definition of Variables*

| Variable Name | Definition |
|---|---|
| Dependent variable | |
| AUTOREF | Number of auto refinishing shops |
| Predictors | |
| URBAN | Percentage of urban land use |
| HWY | Miles of highway |
| POP | Population density (1000 persons/square miles) |
| RETEMP | Retail employment density (1000 employees/square miles) |
| POPRET | POP * RETEMP |

**Results of Model Estimation**

A number of Poisson regression models were estimated. The simplest model has only one predictor, namely retail employment density (RETEMP). The results of the model are presented in Table 3.2, which shows that about 36 percent of the variations in the spatial allocation surrogate is account for by the model. The term CONS in the table is the intercept term in the regression function. Generally, increasing the number of predictors would improve the fit of the model. Several models with more than one predictor were

*Table 3.2 Estimation Results - Model 3.1*

| | | | |
|---|---|---|---|
| Poisson regression | | Number of obs = | 545 |
| | | Model chi2 (1) = | 723.479 |
| Log Likelihood (slopes=0) | = -990.553 | Prob > chi2 = | 0.000 |
| Log Likelihood | = -628.814 | Pseud0 R2 = | 0.3652 |

| Variable | Coefficient | Std. Err | Asy. T ratio | P\|>t\| |
|---|---|---|---|---|
| RETEMP | 7.1489 | 0.2224 | 32.142 | 0.000 |
| CONS | -1.3139 | 0.0801 | -16.397 | 0.000 |

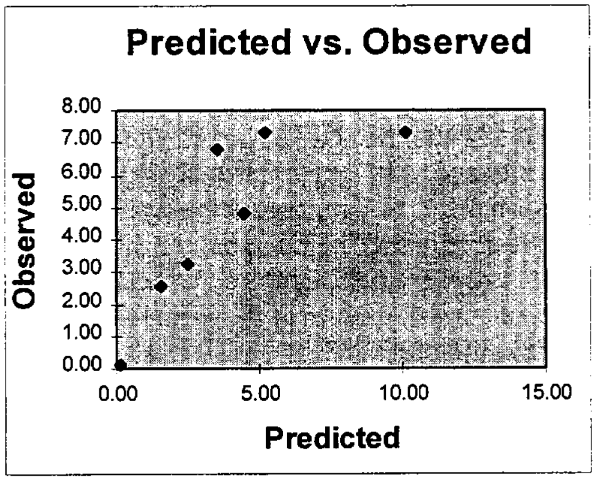
tried, the best among which is presented in Table 3.3. This model accounts for about 60 percent of the variations in the spatial allocation surrogate. Therefore, it is much better than the model shown in Table 3.2. The model in Table 3.3 has five predictors: percentage of urban land use (URBAN), miles of highways (HWY), population density (POP), retail employment density (RETEMP), and an interaction term (POPRET). The first four coefficients have a positive sign, indicating that the number of auto refinishing shops per grid cell is positively correlated to the percentage of urban land use, miles of highways, population density and retail employment density of the grid cell. The coefficient of POPRET has a negative sign, suggesting that POPRET is negatively correlated with the dependent variable. The T test values show that all coefficients are statistically different from zero at the 0.05 significance level. The model in Table 3.3 can be used for updating or predicting the values of the spatial allocation surrogate used for allocating auto refinishing emissions.

***Table 3.3 Estimation Results - Model 3.2***

| | | | |
|---|---|---|---|
| Poisson regression | | Number of obs = | 545 |
| | | Model chi2 (5) = | 1197.479 |
| Log Likelihood (slopes=0) | = -990.553 | Prob > chi2 = | 0.000 |
| Log Likelihood | = -391.779 | Pseud0 R2 = | 0.6045 |

| Variable | Coefficient | Std. Err | Asy. T ratio | P\|>t\| |
|---|---|---|---|---|
| URBAN | 0.0414 | 0.0047 | 8.757 | 0.000 |
| HWY | 0.1462 | 0.0203 | 7.190 | 0.000 |
| POP | 0.3820 | 0.1095 | 3.489 | 0.000 |
| RETEMP | 3.1034 | 1.3478 | 2.303 | 0.022 |
| POPRET | -1.2737 | 0.2223 | -5.728 | 0.000 |
| CONS | -2.6814 | 0.1504 | -17.825 | 0.000 |

*Figure 3.2 Predicted vs. Observed Counts (Auto Refinishing)*




In addition to the statistics in Table 3.3, a scatter plot is given in Figure 3.2 to provide a visual comparison of the predictions with the observed data. Since the count variable has hundreds of values and many of them are zeros (most of the 545 cells do not contain an auto refinishing facility), we need to summarize the data for displaying. One way is to group the data using prediction intervals (intervals selected based on prediction values), compute the group averages of predicted and observed values, and plot them. Seven intervals based on prediction values are used in grouping the data: (0.00 to 0.99), (1.00 to 1.99), (2.00 to 2.99), (3.00 to 3.99), (4.00 to 4.99), (5.00 to 5.99), (6.00 and larger). Seven pairs of groups averages are computed for the predicted and observed values and are plotted in Figure 3.2. On the plot the average prediction values are displayed on the horizontal axis and the average observed counts on the vertical axis. The plot shows the difference between the average predictions and observations. For example, in the prediction interval (2.00 to 2.99), the average predicted count is 2.48, while the average observed count is 3.25. In the interval (4.00 to 4.99), the average prediction is 4.45 and the average observed value is 4.80. The difference is larger of those cells containing 6 or more auto refinishing facilities: the average predicted value is 10.1, whereas the average of observed counts is 7.28. The differences between the average predictions and the average observed values are reasonably small for most prediction intervals. The prediction errors are mainly due to the data used to estimate the model, which are easily available but limited in "prediction power". Similar procedures are used in making the scatter plots for other source categories.

### Effects of Grid Locations on Model Results

In this study the basic spatial unit for allocation and analysis is a 4 km by 4 km model grid cell. The grid system can be placed in many different ways. In other words, the location of the seed cell is chosen more or less randomly. In spatial analysis the effects of a zoning system on results of statistical analysis is called the modifiable areal unit problem (Openshaw, 1984). Here we are concerned with the effects of randomly selecting the location of the seed cell, given cell size, on the results of the regression models, especially the regression coefficients. Our approach to evaluating the effects is to do numerical experiments. The procedure is as follows:

- create a number of grid systems with randomly selected seed cell location
- generate sets of data by spatially overlaying the data layers with the grid systems
- estimate a number of models using the data sets
- examine the sample distribution of the regression coefficients.

Using this procedure we obtained 40 sets of data and estimated 40 Poisson regression models. Table 3.4 presents the distributions of those 40 sets of coefficients. Table 3.4 shows that in general the means are much larger than the standard deviations. The ratio of the mean to its standard deviation ranges roughly from 1 to 15. This suggests that regression coefficients are relatively stable as long as the location of the seed cell is selected randomly. Table 3.5 gives the sample distribution of the standard deviations of the coefficients. The ratio of the mean to its standard deviation ranges approximately from 15 to 32. This shows that the standard deviations are even more stable. It is also noticed that the standard deviation of mean in Table 3.4 is always larger than the mean of standard deviation in Table 3.5. For example, the standard deviation of mean value for URBAN is 0.0069, while the mean of standard deviation for the same coefficient is 0.0047. This shows that the standard deviations of coefficients would become larger as a result of random sampling.

***Table 3.4 Sensitivity Analysis I - Regression Coefficients***

| | CONST | URBAN | HWY | POP | RETEMP | POPRET |
|---|---|---|---|---|---|---|
| MEAN | -2.6854 | 0.0465 | 0.1185 | 0.3590 | 1.5412 | -1.1072 |
| STD | 0.1711 | 0.0069 | 0.0267 | 0.1340 | 1.6753 | 0.3885 |
| RATIO | -15.6980 | 6.7476 | 4.4360 | 2.6786 | 0.9200 | -2.8504 |

*Table 3.5 Sensitivity Analysis II - Standard Deviations of Regression Coefficients*

| | CONST | URBAN | HWY | POP | RETEMP | POPRET |
|---|---|---|---|---|---|---|
| MEAN | 0.1537 | 0.0047 | 0.0205 | 0.1116 | 1.3996 | 0.2479 |
| STD | 0.0101 | 0.0003 | 0.0014 | 0.0053 | 0.0434 | 0.0144 |
| RATIO | 15.2199 | 16.6056 | 14.7682 | 20.9478 | 32.2694 | 17.2410 |

## Temporal Activity Profiles and Ambient Temperature Effects

### The Survey

The auto refinishing inventory survey was conducted during October and November of 1996. The purpose of the survey was to collect data on the temporal activity patterns and ambient temperature effects. A random sample of auto refinishers were selected and telephone interviews were conducted. The survey resulted in 72 completed questionnaires. The spatial distribution of the respondents was 26 in Sacramento county, 26 in Solano county, and 20 in Yolo county.

STI (Sonoma Technology Inc.) designed the survey questionnaire. The questionnaire included questions regarding painting frequencies (day, week, quarter), effects of temperature (90 degrees Fahrenheit and above) and weather (rain) on painting procedures, and some general questions. The questionnaire is presented in Appendix A. Upon completion of the telephone interviews, we designed a coding method for interpreting the results of the survey. We coded the surveys onto an Excel spreadsheet placing the questions asked across the top row and the survey recipients down the far left column.

### Temporal Activity Profiles and Temperature Effects

The temporal activity profiles and temporal allocations are summarized in Tables 3.6-3.10. The survey results show that there is little quarterly variation in auto refinishing activities (see Table 3.6). On average 24.9% of the work was done in the first quarter, 24.8% in the second quarter, 25.3% in the third quarter, and 25% in the fourth quarter. The weekly activity pattern is 95.75% on weekdays, a little on Saturday (3.98%), and almost none on Sunday (0.27%). The daily pattern is almost all work was performed during the daytime with a little bit more done in the morning (51.1%) than in the afternoon (47%). Only 2.3% of the job was done during the evening and none at night. The results also show that temperature and weather have effects on the use of painting materials (see Table 3.10). 84.7% of the respondents reported altering painting

***Table 3.6 Temporal Activity Profiles (Auto Refinishing)***

| CATEGORY | QUARTERLY OR SEASONAL | | | | WEEKLY | | | DIURNAL | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Q1 | Q2 | Q3 | Q4 | Monday to Friday | Saturday | Sunday | 7 a.m.- 12 p.m. | 12p.m.- 6p.m. | 6 p.m. - 0 a.m. |
| Auto Refinishing | 24.9 | 24.8 | 25.3 | 25.0 | 97.5 | 2.4 | 0.1 | 51.0 | 47.0 | 2.0 |

***Table 3.7 Monthly Activity Profiles (Auto Refinishing)***

| CATEGORY | MONTHLY ACTIVITY PROFILE (%) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
| Auto Refinishing | 8.3 | 8.3 | 8.3 | 8.3 | 8.3 | 8.3 | 8.4 | 8.4 | 8.4 | 8.3 | 8.3 | 8.3 |

***Table 3.8 Weekly Activity Profiles (Auto Refinishing)***

| CATEGORY | WEEKLY ACTIVITY PROFILE (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday |
| Auto Refinishing | 19.5 | 19.5 | 19.5 | 19.5 | 19.5 | 2.4 | 0.1 |

***Table 3.9 Diurnal Activity Profiles (Auto Refinishing)***

| CATEGORY | HOURLY ACTIVITY PROFILE (%) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0-1 | 1-2 | 2-3 | 3-4 | 4-5 | 5-6 | 6-7 | 7-8 | 8-9 | 9-10 | 10-11 | 11-12 |
| Auto Refinishing | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 10.2 | 10.2 | 10.2 | 10.2 | 10.2 |
| | 12-13 | 13-14 | 14-15 | 15-16 | 16-17 | 17-18 | 18-19 | 19-20 | 20-21 | 21-22 | 22-23 | 23-24 |
| Auto Refinishing | 7.8 | 7.8 | 7.8 | 7.8 | 7.8 | 7.8 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 |

***Table 3.10 Temperature and Weather Effects (Auto Refinishing)***

| CATEGORY | TEMPERATURE EFFECTS | | | | | WEATHER EFFECTS | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Alter Procedures on Hot Days | On Hot Days | | | | Alter Procedures on Rainy Days | On Rainy Days | | | |
| | | Likely Use in Morning | Likely Use in Evening | Don't Use | Change Formu-lation | | Likely Use in Morning | Likely Use in Evening | Don't use | Change Formu-lation |
| Auto Refinishing | 84.7 | 27.8 | 1.4 | 6.9 | 76.4 | 62.5 | 1.4 | 5.6 | 0.0 | 56.9 |

procedures on hot days. Of the 84.7%, 90.2% changed paint formulation on hot days, and 32.8% were more likely to paint in the morning in order to avoid the afternoon heat. As far as the weather is concerned, 62.5% of the respondents reported change of paint formulation on rainy days.

Monthly allocation factors (in percentage) are shown in Table 3.7. The allocation was made by assuming that the activity is uniformly distributed within a quarter. Weekly allocation factors in percentage are presented in Table 3.8. Uniform distribution of the activity during weekdays is assumed. To obtain the hourly distribution shown in Table 3.8, uniform distributions within the time periods (morning, afternoon, evening) are assumed.

### Confidence Intervals

The confidence intervals computed for the quarterly, weekly, and diurnal allocation estimates, and temperature and weather effect estimates are presented in Table 3.11 to Table 3.14.

***Table 3.11 Confidence Intervals For Quarterly Estimates (Auto Refinishing)***

| Period | Mean | 95% Confidence Interval |
|---|---|---|
| Quarter 1 | 9.4 | [ 7.5, 11.3 ] |
| Quarter 2 | 9.3 | [ 7.4, 11.3 ] |
| Quarter 3 | 9.5 | [ 7.5, 11.5 ] |
| Quarter 4 | 9.4 | [ 7.5, 11.3 ] |

Notes:
(1) Quarter 1 (January, February, March), Quarter 2 (April, May, June),
Quarter 3 (July, August, September), Quarter 4 (October, November, December)
(2) n=72, value = number of vehicles painted per week during the quarter

*Table 3.12 Confidence Intervals For Weekly Estimates (Auto Refinishing)*

| Period | Mean | 95% Confidence Interval |
|---|---|---|
| Monday - Friday | 20.0 | N/A |
| Saturday | 0.5 | [ 0.30, 0.80 ] |
| Sunday | 0.01 | [ 0.00, 0.03 ] |

Notes:
n=72, value = days per month

*Table 3.13 Confidence intervals For Diurnal Estimates (Auto Refinishing)*

| Period | Mean | 95% Confidence Interval |
|---|---|---|
| Morning | 0.511 | [ 0.380, 0.662 ] |
| Afternoon | 0.470 | [ 0.340, 0.618 ] |
| Evening | 0.023 | [ 0.000, 0.163 ] |

Notes:
n=72, value = fraction of vehicles painted on an average day

*Table 3.14 Confidence intervals For Temperature and Weather Effects (Auto Refinishing)*

| | Altering Procedures on Hot Days | Altering Procedures on Rainy Days |
|---|---|---|
| Count | 61 | 45 |
| Relative Frequency | 0.847 | 0.625 |
| Standard Deviation | 0.083 | 0.112 |
| 95% Confidence Intervals | [ 0.764, 0.930 ] | [ 0.513, 0.737 ] |

Notes:
n=72, value = fraction

# 4. Adhesives and Sealants

## Source Activity and Sample Selection

The source category of adhesives and sealants is used by the ARB to inventory the total organic gas (TOG) emissions contained in adhesives and sealants. Applications for adhesives and sealants span almost the entire range of industries (Skeist, 1977). The key industries using adhesives/sealants are, however, limited to a few. A recent study on inventory database for area source solvent emissions (Battye et al., 1993) indicates that the largest users of adhesives and sealants are the paper packaging and wood products industries, which account for over 80 percent of total industrial adhesive and sealant solvent use. Thus, we chose those two industries, plus the furniture and home-building industry and the printing industry, as the sample source industries for this study.

The selected industries are defined by SIC (Standard Industrial Classification) codes 24, 25, 26 and 27 (see Table 4.1 below). Using the SIC codes we assembled a list of businesses in these industries and in the study area. The information was from two sources. The first source was Microcosm (Duns Marketing Service, 1996) - a publication of business lists, available from the Sacramento Chamber of Commerce library. However, the library only had copies for Sacramento County and Yolo County. We purchased the list of businesses by SIC codes in Solano County from the Business Prospector (1996). The lists contained information on firm names, their SIC codes and address (number and street, city, zip code), contact phone number, etc. The combined list included 527 businesses by the SIC codes in the study area. While the bookbinding businesses, large printing houses and publishing companies use significant amounts of adhesives, the small printing shops and publishing offices usually don't use a lot of adhesives. Therefore, the small businesses (less than 15 employees) in SIC code 27 were excluded from the sampling frame. The final list consisted of 312 businesses, of which 225 were in Sacramento County, 54 in Solano County, and 33 in Yolo County. The spatial distribution of selected industrial and commercial users of adhesives and sealants is shown in Figure 4.1. Figure 4.2 provides a comparison of the average predictions with the average actual counts. As previously explained, displayed on the horizontal axis is the average predicted counts and on the vertical axis is the average observed counts. Seven prediction intervals are used in grouping the data: (0.00 to 0.99), (1.00 to 1.99), (2.00 to 2.99), (3.00 to 3.99), (4.00 to 4.99), (5.00 to 5.99), and (6 or larger). The scatter plot shows the differences between the actual and predicted averages, which seem to be reasonably small except the one with the last group (6 or larger).

*Figure 4.1 Spatial Distribution of Major Users of Adhesives and Sealants*

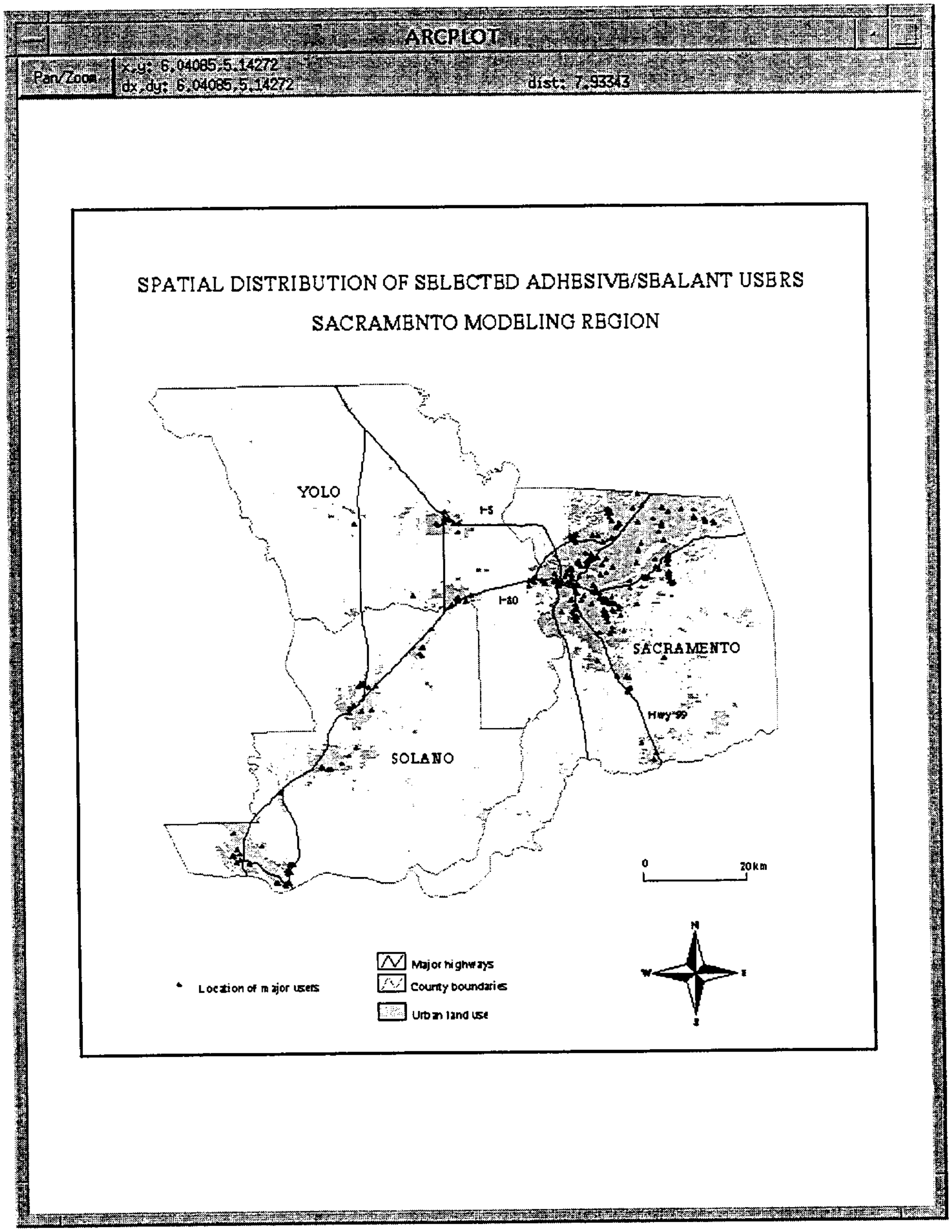


*Figure 4.2 Predicted Counts vs. Observed Counts (Adhesives/Sealants)*

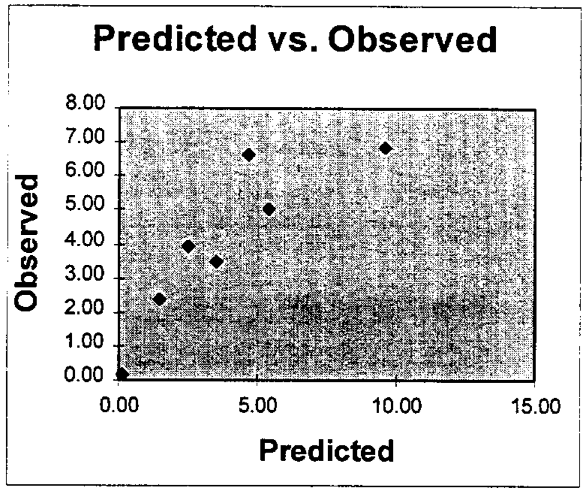



*Table 4.1 Selected Industries by SIC Codes*

| SIC | Industry |
|---|---|
| 24 | Wood products |
| 25 | Furniture and home-building |
| 26 | Packaging |
| 27 | Printing and publishing |

*Table 4.2 Definition of Variables*

| Variable Name | Definition |
|---|---|
| Dependent variable | |
| ADHSLN | Number of major users |
| Predictors | |
| URBAN | Percentage of urban land use |
| HWY | Miles of highway |
| POP | Population density (1000 persons/square miles) |
| MFGEMP | Manufacturing employment density (1000 employees/square miles) |
| POPMFG | POP * MFGEMP |

## Spatial Surrogate for Allocation

### The Spatial Surrogate and Data

The spatial surrogate (spatial activity indicator) selected for this category is the number of businesses that potentially use significant amounts of adhesives and sealants in a model grid cell. The businesses are from the final list. As has been discussed before, a Poisson regression model can be used to estimate or update the spatial allocation surrogate which is a count variable. In the Poisson models for this category the dependent variable (ADHSLN) is the number of adhesives and/or sealants users in a grid cell. The predictors are percentage of urban land use (URBAN), miles of highways (HWY), population density (POP), and manufacturing employment density (MFGEMP). Table 4.2 lists the variables used in the Poisson regression models. The sources of data for the predictors have been described in Section 2. To obtain grid based data, all spatial data layers (location of the users, land use, highways, population census) were overlaid with the model grid layer. Using the AML programs, grid based data values were computed. The data were exported from the GIS database and then were read into a statistical program. Poisson regression models were then estimated.

### Results of Model Estimation

A number of Poisson regression models were estimated using different specifications. The model, shown in Table 4.3 has only one predictor - manufacturing employment density (MFGEMP). This model accounts for about 25 percent of the variations in the dependent variable. A number of alternative models estimated, among which the model presented in Table 4.4 fitted the data best, account for about 51 percent of the variations. The model in Table 4.4 has five predictors: percentage of urban land use (URBAN), miles of highways (HWY), population density (POP), manufacturing employment density (MFGEMP), and an interaction term (POPMFG). All coefficients are statistically different from zero. Most regression coefficients have expected signs. For example, it is expected that most industrial and commercial users of adhesives and sealants are located in urban areas, near highways or major roads, or in areas with relatively high manufacturing employment. Therefore, the positive sign of the coefficients of URBAN, HWY, MFGEMP is consistent with our expectation. The negative sign of the coefficient of the population density variable has dubious meanings. It might reflect the fact that businesses in some of the selected industries, such as wood and home-building products, tend to be located away from the population center (city center). Or it could be the result of multicollinearity (correlation between the regressors), which we will discuss next. The model in Table 4.4 fits the data reasonably well and is easy to use. Thus, this model is recommended for this category.

*Table 4.3 Estimation Results - Model 4.1*

| | | | |
|---|---|---|---|
| Poisson regression | | Number of obs = | 545 |
| | | Model chi2 (1) = | 442.937 |
| Log Likelihood (slopes=0) | = -888.086 | Prob > chi2 = | 0.000 |
| Log Likelihood | = -666.618 | Pseud0 R2 = | 0.249 |

| Variable | Coefficient | Std. Err | Asy. T ratio | P\|>t\| |
|---|---|---|---|---|
| MFGEMP | 11.9018 | 0.4343 | 27.404 | 0.000 |
| CONS | -1.1599 | 0.0741 | -15.644 | 0.000 |

*Table 4.4 Estimation Results - Model 4.2*

| | | | |
|---|---|---|---|
| Poisson regression | | Number of obs = | 545 |
| | | Model chi2 (5) = | 902.729 |
| Log Likelihood (slopes=0) | = -888.086 | Prob > chi2 = | 0.000 |
| Log Likelihood | = -436.722 | Pseud0 R2 = | 0.508 |

| Variable | Coefficient | Std. Err | Asy. T ratio | P\|>t\| |
|---|---|---|---|---|
| URBAN | 0.0554 | 0.0047 | 11.886 | 0.000 |
| HWY | 0.1725 | 0.0209 | 8.254 | 0.000 |
| POP | -0.3560 | 0.1200 | -2.966 | 0.003 |
| MFGEMP | 9.4942 | 2.3469 | 4.045 | 0.000 |
| POPMFG | -1.6249 | 0.4575 | -3.552 | 0.000 |
| CONS | -2.3794 | 0.1335 | -17.818 | 0.000 |

The model in Table 4.4 has a problem - several of the independent variables (URBAN, POP, MFGEMP) in the regression equation are highly correlated. This problem is known as multicollinearity and may result in less accurate estimates and/or wrong signs of the regression coefficients. Several methods are available for coping with the problem of multicollinearity. The obvious remedy is to drop the variables causing the problem from the regression. In doing so, however, one loses valuable information and may encounter the problem of specification. A better approach is to use principal components (e.g., Johnson and Wichern, 1992) that are linear combinations of the original correlated variables. The objective of principal component analysis is to construct uncorrelated linear combinations of the variables that account for much of the variations in the data. The sample principal components are those linear combinations which have maximum sample variance. The procedure for using principal components in the regression is straightforward. First, the linear combinations of the strongly correlated variables are computed. The principal components are then selected based on the eigen-values. Finally, the dependent variable is regressed on the principal components and any remaining uncorrelated predictors.

***Table 4.5 Principal Component Analysis***

| | EIGENVECTORS | | |
|---|---|---|---|
| Variable | PC1 | PC2 | PC3 |
| 1. URBAN | 0.5722 | 0.7412 | 0.3510 |
| 2. POP | 0.5851 | -0.0691 | -0.8080 |
| 3. MFGEMP | 0.5747 | -0.6677 | 0.4732 |
| Cumulative proportion of total (standardized) sample variation explained | 0.9513 | 0.9885 | 1.0000 |

*Table 4.6 Estimation Results - Model 4.3*

| | | | | |
|---|---|---|---|---|
| Poisson regression | | | Number of obs = | 545 |
| | | | Model chi2 (5) = | 809.735 |
| Log Likelihood (slopes=0) | = -888.086 | | Prob > chi2 = | 0.000 |
| Log Likelihood | = -483.219 | | Pseud0 R2 = | 0.456 |

| Variable | Coefficient | Std. Err | Asy. T ratio | P\|>t\| |
|---|---|---|---|---|
| PC1 | 0.4199 | 0.0193 | 21.720 | 0.000 |
| PC2 | 0.9928 | 0.0834 | 11.904 | 0.000 |
| HWY | 0.1780 | 0.0198 | 9.014 | 0.000 |
| CONS | -1.7473 | 0.1068 | -16.365 | 0.000 |

Table 4.5 presents the results of the principal component analysis. 95% of the variation in the three variables (URBAN, POP, MFGEMP) is explained by the first principal component (PC1). The first two components (PC1 and PC2) together account for almost 99% of the variation. The first principal component is a measure of the overall urban environment since the first eigenvector shows approximately equal loadings on all three variables. The second eigenvector has high positive loadings on URBAN, and negative loadings on POP and MFGEMP. It appears that the second component contrasts the urban land use variable with the population and employment density variables. The third component, PC3, has large negative loadings on POP and positive loadings on URBAN and MFGEMP. Thus, the second and third components seem to measure a tendency: that some of the industries selected for the study are more likely to be located at urban peripheries where population density is relatively low.

The first two principal components plus HWY are the predictors in the model presented in Table 4.6. All regression coefficients of the model are statistically significantly different from zero. The R-square value is 0.456, which is smaller than that of the previous model. This suggests the model fits the data less well than does the model in Table 4.4. The use of principal components in the regression model avoid the multicollinearity problem. However, it makes computations more complicated and the interpretation of regression results less straightforward.

### Effects of Grid Locations on Regression Results

In this study the unit of spatial modeling is a grid cell. The origin of the spatial grid system is usually chosen arbitrarily. The question is whether or not the choice of grid locations, given cell size, would have considerable effects on the regression coefficients. As discussed in Section 3, we use an empirical approach to examine the sensitivity of grid location on regression results. The process consisted of the following steps.

(1) 45 grid systems with same cell size (4 km by 4 km) but random origins (seed cell locations) were created.

(2) The grid layers were integrated with the regression data layers to generate 45 estimation data sets.

(3) Using each of the 45 data sets, two Poisson regression models were estimated. One used the specification of the model in Table 4.4; the other used the variable specification in Table 4.6 with principal components.

The results of analyses are presented in Table 4.7 and Table 4.8. The tables give the sample distributions of the two sets of regression coefficients - their means, standard deviations, and the ratio of a mean to its standard deviation. The sample size is 45. Note that the smaller the standard deviations, or the larger the ratios of the means to the standard deviations, the more stable the regression results. Table 4.7 shows the results of the regression models without principal components. The ratio of CONST, URBAN, and HWY are large ranging from 5 to 16 in absolute value. The ratios of POP, MFGEMP, and POPMFG are relatively small, less than 2 in absolute value. It is suggested that some of the regression coefficients are quite stable, while the others are less so. The results of using principal components are better. Table 4.8 shows that the ratio of all coefficients estimated using principal components range from 4.8 to 25.1 in absolute value.

***Table 4.7 Sensitivity Analysis I - Regression without principal components***

| | CONST | URBAN | HWY | POP | MFGEMP | POPMFG |
|---|---|---|---|---|---|---|
| MEAN | -2.3941 | 0.0557 | 0.1501 | -0.2334 | 9.5743 | -2.3314 |
| STD | 0.1474 | 0.0109 | 0.0229 | 0.2570 | 5.3765 | 1.3828 |
| RATIO | -16.2428 | 5.1175 | 6.5594 | -0.9082 | 1.7808 | -1.6860 |

*Table 4.8 Sensitivity Analysis II - Regression with principal components*

| | CONST | PC1 | PC2 | HWY |
|---|---|---|---|---|
| MEAN | -1.7585 | 0.4141 | 1.1178 | 0.1581 |
| STD | 0.0698 | 0.0324 | 0.2321 | 0.0216 |
| RATIO | -25.1807 | 12.7644 | 4.8163 | 7.3234 |

## Temporal Activity Profiles and Ambient Temperature Effects

### The Survey

The survey for the adhesives/sealants category was conducted during January and February of 1997. The purpose of the survey was to collect data on temporal activity patterns and ambient temperature effects. A random sample of 90 was initially selected for the survey. The sample size was later increased to 156 due to insufficient responses. The number of surveys completed was 68. The spatial distribution of the respondents was 31 in Sacramento county, 20 in Solano county, and 17 in Yolo county.

The questionnaire from the previous survey (auto refinishing) was modified to be used in this survey. The questionnaire included questions regarding temporal activities (day, week, quarter), effects of temperature (90 degrees Fahrenheit and above) and weather (rain) on application procedures, and some general questions. The questionnaire is shown in the Appendix. The results of the survey were coded and entered onto an Excel spreadsheet.

The completion rate of the survey was relatively low (68/156=44%). There were at least three reasons for this: (1) The survey subjects were identified by SIC codes, not by actual use of adhesives/sealants. Although the selected industries in general may consume large amounts of adhesives/sealants, not every company in the industry uses significant amounts. Some don't even consider themselves the users of adhesives/sealants. The limitation of using SIC codes to identify survey objects was a factor. (2) Unlike the auto refinishing surveys in which everyone we called knew that auto refinishing was one of their services and who specifically was providing this service, many people we contacted for the adhesives/sealants surveys were not sure if their companies had the materials and who might be using them. Usually the larger the company, the more difficult to find the right person to answer the questions on our surveys. (3) Some people just were uncooperative.

### Temporal Activity Profiles and Temperature Effects

The temporal activity profiles and temporal allocations are summarized in Tables 4.9-4.13. The survey results show that there is little quarterly variations in the source activities. Table 4.9 shows that on average 24.9% of the usage was in the first quarter, 24.9% in the second quarter, 25.0% in the third quarter, and 25.2% in the fourth quarter. The weekly activity pattern is 92.8% on weekdays, a little on Saturday (5.6%) and Sunday (1.6%). The daily pattern is 50% in the morning, 49% in the afternoon, and 1% in the evening. Table 4.13 shows that the effects of ambient temperature and weather on the use of adhesives and sealants are small: 10.3% of the respondents would alter application procedures on hot days, 14.7% would do so on rainy days.

Monthly allocation factors (in percentage) are presented in Table 4.10. The allocation was made by assuming that the activity is uniformly distributed within a quarter. Weekly allocation factors in percentage are provided in Table 4.11. It is assumed that the activities during weekdays are homogenous. To obtain the hourly distribution shown in Table 4.12, uniform distributions within the time periods (morning, afternoon, evening) are assumed.

### Confidence Intervals

The confidence intervals for the estimates on quarterly, weekly, and diurnal allocations are presented in Table 4.14 to 4.16. The confidence intervals for estimates of temperature and weather effects estimates are given in Table 4.17.

*Table 4.9 Temporal Activity Profiles (Adhesives/Sealants)*

| CATEGORY | QUARTERLY OR SEASONAL | | | | WEEKLY | | | DIURNAL | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Q1 | Q2 | Q3 | Q4 | Monday to Friday | Saturday | Sunday | 7 a.m.- 12 p.m. | 12p.m.- 6p.m. | 6 p.m. - 0 a.m. |
| Adhesives/Sealants | 24.9 | 24.9 | 25.0 | 25.2 | 92.8 | 5.6 | 1.6 | 50.0 | 49.0 | 1.0 |

*Table 4.10 Monthly Activity Profiles (Adhesives/Sealants)*

| CATEGORY | MONTHLY ACTIVITY PROFILE (%) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
| Adhesives/Sealants | 8.3 | 8.3 | 8.3 | 8.3 | 8.3 | 8.3 | 8.3 | 8.3 | 8.3 | 8.4 | 8.4 | 8.4 |

***Table 4.11 Weekly Activity Profiles (Adhesives/Sealants)***

| CATEGORY | WEEKLY ACTIVITY PROFILE (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday |
| Adhesives/Sealants | 18.6 | 18.6 | 18.6 | 18.6 | 18.6 | 5.6 | 1.6 |

***Table 4.12 Diurnal Activity Profiles (Adhesives/Sealants)***

| CATEGORY | HOURLY ACTIVITY PROFILE (%) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0-1 | 1-2 | 2-3 | 3-4 | 4-5 | 5-6 | 6-7 | 7-8 | 8-9 | 9-10 | 10-11 | 11-12 |
| Adhesives/Sealants | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 |
| | 12-13 | 13-14 | 14-15 | 15-16 | 16-17 | 17-18 | 18-19 | 19-20 | 20-21 | 21-22 | 22-23 | 23-24 |
| Adhesives/Sealants | 8.2 | 8.2 | 8.2 | 8.2 | 8.2 | 8.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |

***Table 4.13 Temperature and Weather Effects (Adhesives/Sealants)***

| CATEGORY | TEMPERATURE EFFECTS | | | | | WEATHER EFFECTS | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Alter Procedures on Hot Days | On Hot Days | | | | Alter Procedures on Rainy Days | On Rainy Days | | | |
| | | Likely Use in Morning | Likely Use in Evening | Don't Use | Change Formu-lation | | Likely Use in Morning | Likely Use in Evening | Don't use | Change Formu-lation |
| Adhesives/Sealants | 10.3 | 1.5 | 0.0 | 1.5 | 7.4 | 14.7 | 1.5 | 4.4 | 0.0 | 8.8 |

*Table 4.14 Confidence Intervals For Quarterly Estimates (Adhesives/Sealants)*

| Period | Mean | 95% Confidence Interval |
|---|---|---|
| Quarter 1 | 24.9 | [ 24.7, 25.0 ] |
| Quarter 2 | 24.9 | [ 24.8, 25.0 ] |
| Quarter 3 | 25.0 | [ 24.9, 25.2 ] |
| Quarter 4 | 25.2 | [ 24.8, 25.5 ] |

Notes:
(1) Quarter 1 (January, February, March), Quarter 2 (April, May, June),
Quarter 3 (July, August, September), Quarter 4 (October, November, December)
(2) n=66, value = percentage

*Table 4.15 Confidence Intervals For Weekly Estimates (Adhesives/Sealants)*

| Period | Mean | 95% Confidence Interval |
|---|---|---|
| Monday - Friday | 20.0 | N/A |
| Saturday | 1.21 | [ 0.87, 1.54 ] |
| Sunday | 0.34 | [ 0.10, 0.58 ] |

Notes:
n=68, value = days per month

*Table 4.16 Confidence Intervals For Diurnal Estimates (Adhesives/Sealants)*

| Period | Mean | 95% Confidence Interval |
|---|---|---|
| Morning | 0.505 | [ 0.470, 0.540 ] |
| Afternoon | 0.495 | [ 0.460, 0.530 ] |
| Evening | 0.015 | [ 0.000, 0.120 ] |

Notes:
n=68, value = fraction of adhesives/sealants used on an average day

*Table 4.17 Confidence Intervals For Temperature and Weather Effects (Adhesives/Sealants)*

| | Altering Procedures on Hot Days | Altering Procedures on Rainy Days |
|---|---|---|
| Count | 7 | 10 |
| Relative Frequency | 0.103 | 0.147 |
| Standard Deviation | 0.072 | 0.043 |
| 95% Confidence Intervals | [ 0.031, 0.175] | [ 0.063, 0.231 ] |

Notes:
n=68, value = fraction

# 5. Can & Coil, Metal Parts & Products Coatings

## Source Activity and Sample Selection

This combined category consists of two original source emission categories, can and coil coatings, and metal surface and products coatings, which are used by the ARB to inventory the total organic gas (TOG) emissions from the application of surface coatings in the manufacturing and industrial sectors. Since the coating applications could take place in a broad array of industries, the first step was to identify the industries that were most likely to be involved. The ARB provided us a list of companies and/or facilities in this category from its point source emission inventory database. With the help of the ARB list and other references (e.g., Battye, 1993) we selected several industries that were thought to be most relevant to the surface coating applications. Those industries and their SIC codes are shown in Table 5.1.

***Table 5.1 Selected Industries by SIC Codes***

| SIC | Industry |
|---|---|
| 34 | Fabricated Metal Products (including can, coil, and metal coating) |
| 35 | Industrial Machinery and Equipment (e.g. construction machinery) |
| 37 | Transportation Equipment (e.g. motor vehicles parts and accessories) |
| 2032 | Canned Specialties |
| 2033 | Canned Fruits and Vegetables |

Using the SIC codes we obtained a list of businesses in these industries and in the study area from the Business Prospector. The list contained 452 companies and included information on firm names, their SIC codes and address (number and street, city, zip code), and contact phone number. The next step was to contact the businesses on the list to find out which were actually involved in the coating operations. To work efficiently, we decided to call the companies in Solano county first. Solano county was chosen for the experiment because it is representative of the three-county areas: it is not as industrialized as Sacramento county but it is more urbanized than Yolo county. All ninety one businesses located in the county on the list were called and 90 percent of them were successfully contacted. An overwhelming majority of the businesses contacted told us that either they didn't use metal surface coatings, or that they sent out their products to specialized coating companies. The reason for having products coated by specialized coating businesses is because coatings require special equipment. It is more economical to have coatings done by the specialized companies. It was also found that most of the specialized coating facilities were located in heavily industrialized urban areas. For example, many of the businesses in Solano county sent their products to the San Francisco Bay Area to be coated, while the others sent their products to the Sacramento area or local coating facilities. By SIC codes most coating facilities were found in the 34 category, especially 3479 (metal surface and allied services).

The results of the Solano experiment suggested that we ought to concentrate on the specialized coating facilities. Thus, we further searched local phone books and on-line databases for metal surface coating facilities. In addition we contacted those companies in Sacramento and Yolo counties that were in certain SIC codes shown positive in the Solano experiment. The outcome was that 44 companies or facilities were identified as having coating operations in the study area. The spatial distribution is 24 in Sacramento county, 15 in Solano county, and 5 in Yolo county. Of the 44 observations, 22 were on the ARB list and the rest came out of our search. It should be noted that those 44 facilities represent a sample of the population of interest. We couldn't identify all companies that are involved in metal surface coating applications. The spatial distribution of the coating facilities is shown in Figure 5.1. Figure 5.2 provides a comparison of the average predicted counts with the average actual counts.

## Spatial Surrogate for Allocation

### The Spatial Surrogate and Data

The spatial surrogate (spatial activity indicator) selected for this category is the number of businesses that use metal surface coatings in a model grid cell. The businesses are identified using the ARB list and results of our search. As has been discussed before, a Poisson regression model can be used to estimate or update the spatial allocation surrogate when it is a count variable. In the Poisson models for this category the dependent variable (MSCF) is the number of identified metal surface coating facilities in a grid cell. The predictors are percentage of urban land use (URBAN), miles of highways (HWY), population density (POP), and manufacturing employment density (MFGEMP). Table 5.2 lists the variables used in the Poisson regression models. The sources of data for the predictors have been described in Section 2. To obtain grid based data, all spatial data layers (location of the facilities, land use, highways, population census) were overlaid with the model grid layer. Using the AML programs, grid based data values were computed. The data were exported from the GIS database and then were read into a statistical program. Poisson regression models were then estimated. The models were estimated using the maximum likelihood estimation method.

*Figure 5.1 Spatial Distribution of Identified Metal Surface Coating Facilities*

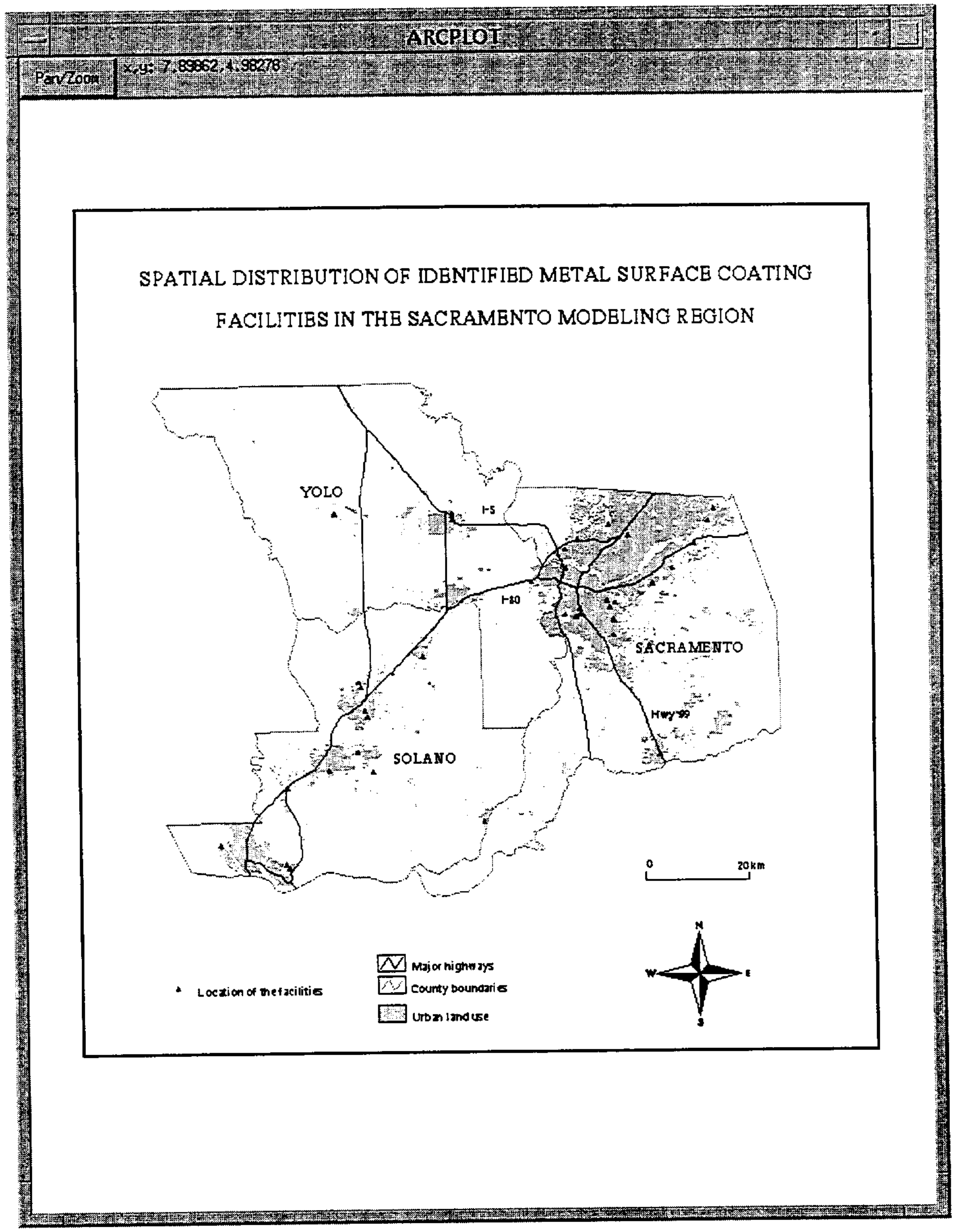


*Figure 5.2 Predicted Counts vs. Observed Counts (Metal Surface Coatings)*

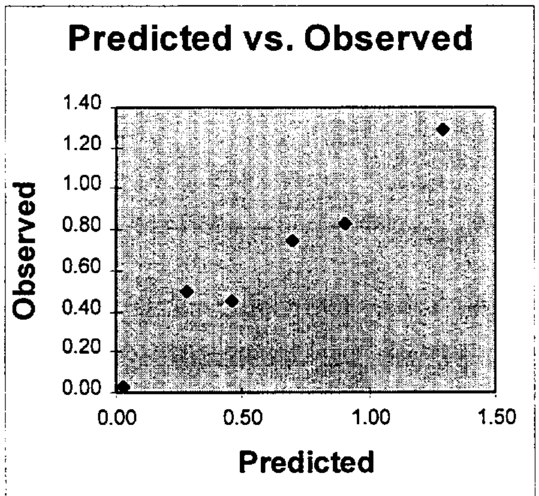



*Table 5.2 Definition of Variables*

| Variable Name | Definition |
|---|---|
| Dependent variable | |
| MSCF | Metal surface coating facilities |
| Predictors | |
| URBAN | Percentage of urban land use |
| HWY | Miles of highway |
| POP | Population density (1000 persons/square miles) |
| MFGEMP | Manufacturing employment density (1000 employees/square miles) |
| POPMFG | POP * MFGEMP |

**Results of Model Estimation**

A number of Poisson regression models were estimated using different variable specifications. Table 5.3 shows that only about 15 percent of the variations in the dependent variable is accounted for by the model if a single predictor, manufacturing employment density (MFGEMP) is used. A number of alternative models with more predictors are estimated, among which the model presented in Table 5.4 fits the data best.

*Table 5.3 Estimation Results - Model 5.1*

| | | | |
|---|---|---|---|
| Poisson regression | | Number of obs = | 545 |
| | | Model chi2 (1) = | 48.262 |
| Log Likelihood (slopes=0) | = -166.343 | Prob > chi2 = | 0.000 |
| Log Likelihood | = -142.213 | Pseud0 R2 = | 0.1451 |

| Variable | Coefficient | Std. Err | Asy. T ratio | P\|>t\| |
|---|---|---|---|---|
| MFGEMP | 11.0255 | 1.2261 | 8.993 | 0.000 |
| CONS | -3.0284 | 0.1914 | -15.824 | 0.000 |

*Table 5.4 Estimation Results - Model 5.2*

| | | | |
|---|---|---|---|
| Poisson regression | | Number of obs = | 545 |
| | | Model chi2 (5) = | 109.66 |
| Log Likelihood (slopes=0) | = -166.344 | Prob > chi2 = | 0.000 |
| Log Likelihood | = -111.512 | Pseud0 R2 = | 0.3296 |

| Variable | Coefficient | Std. Err | Asy. T ratio | P\|>t\| |
|---|---|---|---|---|
| URBAN | 0.0340 | 0.0074 | 4.580 | 0.000 |
| HWY | 0.0867 | 0.0626 | 1.385 | 0.173 |
| MFGEMP | 30.1277 | 6.2730 | 4.803 | 0.000 |
| POPMFG | -6.4914 | 1.4072 | -4.613 | 0.000 |
| CONS | -4.1106 | 0.3273 | -12.561 | 0.000 |

The model in Table 5.4 accounts for about 33 percent of the variations in the dependent variable. The model has four predictors: percentage of urban land use (URBAN), miles of highways (HWY), manufacturing employment density (MFGEMP), and an interaction term (POPMFG). All coefficients are statistically different from zero. The regression coefficients have expected signs. For example, it is expected that most metal surface coating facilities would be located in urban areas, near highways or major roads, or in areas with relatively high manufacturing employment. Therefore, the positive sign of the coefficients of URBAN, HWY, MFGEMP is consistent with our expectation. The regression results are also consistent with the activity pattern shown on the map in Figure 5.1. Since the number of sample facilities in this category is quite small (44 in total), the average count per cell is very small. Thus six small prediction intervals are used to group the data for plotting: (0.0 to 0.2), (0.2 to 0.4), (0.4 to 0.6), (0.6 to 0.8), (0.8 to 1.0), and (1.0 or larger). The plot shows that the differences between the average predictions and the average actual values in the intervals are reasonably small.

## Temporal Activity Profiles and Ambient Temperature Effects

### The Survey

The survey for this source category was conducted during March and April of 1997. The purpose of the survey was to collect data on temporal activity patterns and ambient temperature effects. As discussed in the sample selection section, 44 companies or facilities in the study area were identified to be involved in metal surface coatings. Out of the 44, we were able to complete 22 surveys. Several factors contributed to the low completion rate: (1) some of the survey subjects were not cooperative, (2) some of the facilities on the ARB list could not be surveyed (e.g., military bases), and (3) some of the companies on the ARB list didn't answer our survey questions by saying that they didn't use surface coatings. The spatial distribution of the 22 respondents is 12 in Sacramento county, 9 in Solano county, and 1 in Yolo county. The distribution by coating type is 41% solvent based, 26% water based, 22% power coating, 4% exempt solvent, and 7% high solid.

References (CARB, 1991a, 1991b) and previous questionnaires were used to design the questionnaire for this survey. The questionnaire included questions regarding temporal activities (day, week, quarter), effects of temperature (90 degrees Fahrenheit and above) and weather (rain) on application procedures, and some general questions. The questionnaire is shown in the Appendix. The results of the survey were coded and entered onto an Excel spreadsheet for analysis.

### Temporal Activity Profiles and Temperature Effects

The temporal activity profiles and temporal allocations are summarized in Tables 5.5-5.8. The survey results indicate that there are little quarterly variations in metal surface coating activities as reported by the respondents. Table 5.5 shows that on average 24.6% of the usage was in the first quarter, 25.9% in the second quarter, 25.6% in the third quarter, and 23.9% in the fourth quarter. The weekly pattern was 89.1% of the activity took place Monday-Friday, some on Saturday (8.3%), and a little on Sunday (2.6%). The diurnal pattern was that 46.8% were in the morning, 44.2% in the afternoon, and 8.9% in the evening. Monthly allocation factors in percentage are presented in Table 5.6. The allocation was made by assuming that the activity is uniformly distributed within a quarter. Weekly allocation factors (in percentage) are provided in Table 5.7. It is assumed that the activities during weekdays are homogenous. The hourly distribution shown in Table 5.8 is based on work schedules reported by the respondents. Table 5.9 shows that the effect of ambient temperature on the application of metal surface coatings is relatively small - 18.2% of the respondents would alter application procedures on hot days, while the effect of raining is large - 40.9% of the respondents would alter the procedure on rainy days.

### Confidence Intervals

The confidence intervals for the estimates on quarterly, weekly, and diurnal allocations are presented in Tables 5.10 to 5.12. The confidence intervals for estimates of temperature and weather effects estimates are given in Table 5.13. Notice that the sample is not truly random and the sample size is relatively small. The confidence intervals may not be accurate.

*Table 5.5 Temporal Activity Profiles (Metal Surface Coatings)*

| CATEGORY | QUARTERLY OR SEASONAL | | | | WEEKLY | | | DIURNAL | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Q1 | Q2 | Q3 | Q4 | Monday to Friday | Saturday | Sunday | 6 a.m.- 12 p.m. | 12p.m.- 6p.m. | 6 p.m. - 0 a.m. |
| Can & Coil / Metal Parts and Products Coatings | 24.6 | 25.9 | 25.6 | 23.9 | 89.1 | 8.3 | 2.6 | 46.8 | 44.2 | 8.9 |

*Table 5.6 Monthly Activity Profiles (Metal Surface Coatings)*

| CATEGORY | MONTHLY ACTIVITY PROFILE (%) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
| Can & Coil / Metal Parts and Products Coatings | 8.2 | 8.2 | 8.2 | 8.6 | 8.6 | 8.6 | 8.5 | 8.5 | 8.5 | 8.0 | 8.0 | 8.0 |

*Table 5.7 Weekly Activity Profiles (Metal Surface Coatings)*

| CATEGORY | WEEKLY ACTIVITY PROFILE (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday |
| Can & Coil / Metal Parts and Products Coatings | 17.8 | 17.8 | 17.8 | 17.8 | 17.8 | 8.3 | 2.6 |

*Table 5.8 Diurnal Activity Profiles (Metal Surface Coatings)*

| CATEGORY | HOURLY ACTIVITY PROFILE (%) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0-1 | 1-2 | 2-3 | 3-4 | 4-5 | 5-6 | 6-7 | 7-8 | 8-9 | 9-10 | 10-11 | 11-12 |
| Can & Coil / Metal Parts and Products Coatings | 0.4 | 0.4 | 0.4 | 0.4 | 0.9 | 0.9 | 1.9 | 2.6 | 9.4 | 9.4 | 9.4 | 9.4 |
| | 12-13 | 13-14 | 14-15 | 15-16 | 16-17 | 17-18 | 18-19 | 19-20 | 20-21 | 21-22 | 22-23 | 23-24 |
| Can & Coil / Metal Parts and Products Coatings | 9.4 | 9.4 | 9.2 | 9.0 | 8.8 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 0.9 |

*Table 5.9 Temperature and Weather Effects (Metal Surface Coatings)*

| CATEGORY | TEMPERATURE EFFECTS | | | | | WEATHER EFFECTS | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Alter Procedures on Hot Days | On Hot Days | | | | Alter Procedures on Rainy Days | On Rainy Days | | | |
| | | Likely Use in Morning | Likely Use in Evening | Don't Use | Change Formu-lation | | Likely Use in Morning | Likely Use in Evening | Don't use | Change Formu-lation |
| Can & Coil / Metal Parts and Products Coatings | 18.2 | 4.5 | 4.5 | 0.0 | 13.6 | 40.9 | 0.0 | 0.0 | 18.2 | 18.2 |

*Table 5.10 Confidence intervals For Quarterly Estimates (Metal Surface Coatings)*

| Period | Mean | 95% Confidence Interval |
|---|---|---|
| Quarter 1 | 24.6 | [ 23.3, 25.8 ] |
| Quarter 2 | 25.9 | [ 24.1, 27.8 ] |
| Quarter 3 | 25.6 | [ 23.8, 27.4 ] |
| Quarter 4 | 23.9 | [ 21.5, 26.3 ] |

Notes:
(1) Quarter 1 (January, February, March), Quarter 2 (April, May, June), Quarter 3 (July, August, September), Quarter 4 (October, November, December)
(2) n=22, value = percentage

*Table 5.11 Confidence intervals For Weekly Estimates (Metal Surface Coatings)*

| Period | Mean | 95% Confidence Interval |
|---|---|---|
| Monday - Friday | 20.0 | N/A |
| Saturday | 1.86 | [ 1.11, 2.62 ] |
| Sunday | 0.59 | [ 0.00, 1.18 ] |

Notes:
n=22, value = days per month

*Table 5.12 Confidence intervals For Diurnal Estimates (Metal Surface Coatings)*

| Period | Mean | 95% Confidence Interval |
|---|---|---|
| Morning | 4.50 | [ 4.14, 4.86 ] |
| Afternoon | 4.25 | [ 3.89, 4.61 ] |
| Evening | 0.86 | [ 0.02, 1.70 ] |

Notes:
n=22, value = working hours on an average day

*Table 5.13 Confidence intervals For Temperature and Weather Effects (Metal Surface Coatings)*

| | Altering Procedures on Hot Days | Altering Procedures on Rainy Days |
|---|---|---|
| Count | 4 | 9 |
| Relative Frequency | 0.180 | 0.410 |
| Standard Deviation | 0.160 | 0.210 |
| 95% Confidence Intervals | [ 0.020, 0.340 ] | [ 0.200, 0.620 ] |

Notes:
n=22

# 6. Farm Equipment

The farm equipment category is used by the ARB to inventory the combustion emissions (TOG, CO, NOx, SOx, $PM_{10}$) from the use of tractors, combines, balers, mowers, and other equipment in agriculture production. The emission allocation surrogate chosen for this category is the farm equipment operation time (hours) by crops for a given spatial unit (4 km by 4km grid cell). The machinery operation time required to produce a particular crop type is a critical factor in estimating farm equipment emissions (KVB, 1980; Sierra Research, 1993) and can be computed from sample production cost estimates that are prepared by county farm advisors and the University of California Cooperative Extension (1996). Those estimates are developed to help local farmers to select crops to produce and to provide a basis for farm loans. Another important component in calculating the allocation factors is crop acreage since farm equipment emissions are a function of a crop type and its acreage. Information on crop types and their spatial distribution in the study area is contained in the land use GIS data provided by the California Department of Water Resources. Figure 6.1 shows the spatial distribution of major crops in the study area.

The sample production cost tables also provide information on temporal (monthly) distributions of farm equipment uses. Table 6.1 gives a sample summary of production cost estimates for field corn in Yolo county, California. The table lists the operations performed in producing the crop, the time required for each operation (hours/acre), and the relative activity by month. From the table, monthly activity estimates for the crop can be made. For example, the production cost estimate for the month of October is given by

$$\text{Cost Estimate (October)} = 0.20 + 0.20 + 0.34 + 0.28 = 1.02 \text{ (Hours/Acre)} \qquad (6.1)$$

The result of multiplying 1.02 to the total acreage of corn in a spatial unit is the cost of producing corn for the spatial unit in October. Notice that, in the calculations, we include only those operations involving machinery, so that the result is hours of mechanical work required.

Since sample production cost estimates could not be obtained for all crops in the study area, only major crops were used in the allocation. We have also been advised (Livingston, 1996) that agricultural practices for a crop in one county are usually very similar to those for a neighboring county (e.g., Yolo County vs. Sacramento County) or even a neighboring valley (e.g., the Sacramento Valley vs. the San Joaquin Valley).

*Figure 6.1 Agriculture Land Use Map of the Study Area*

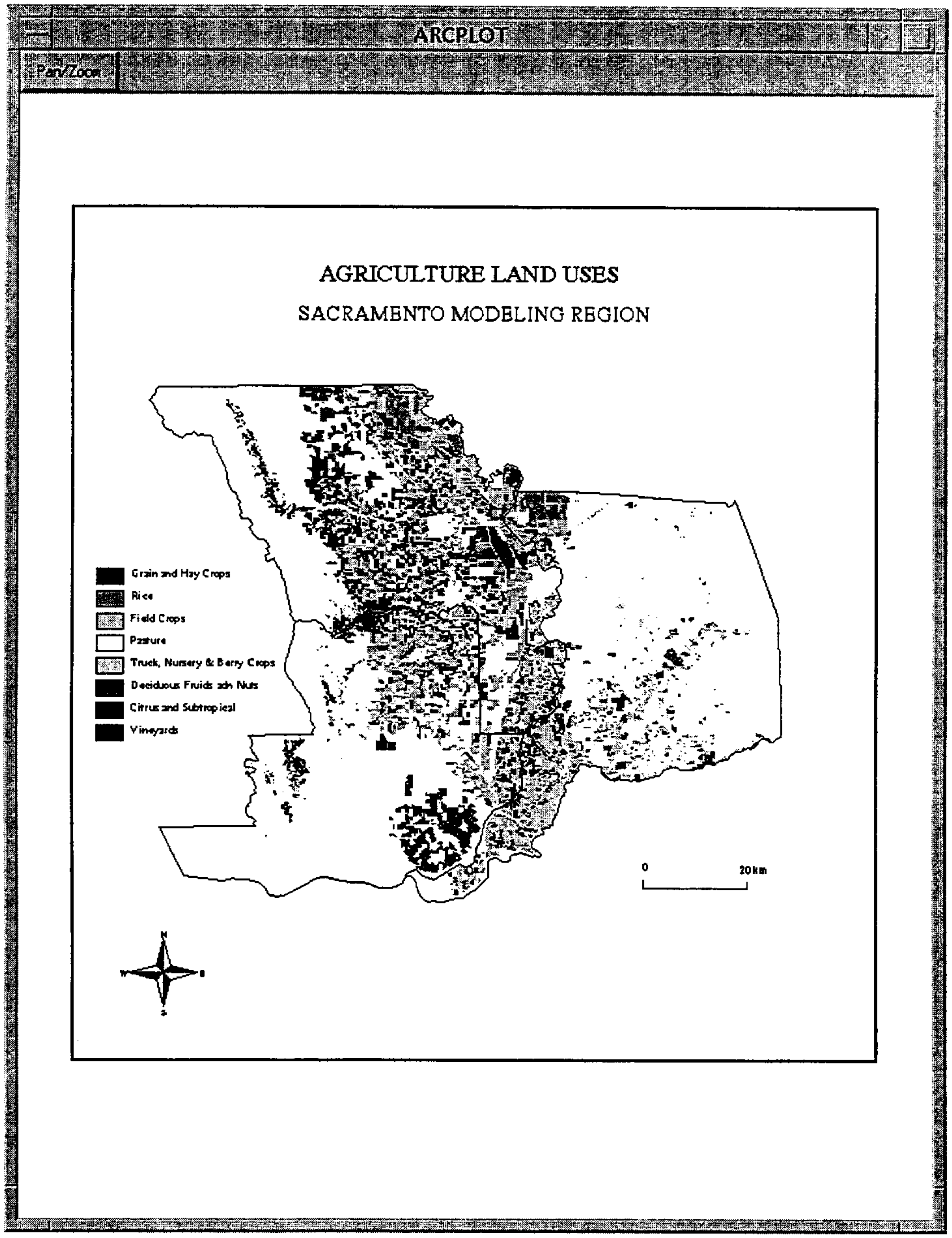


***Table 6.1 Summary of Production Cost Estimate for Field Corn***

| Field Corn: Yolo County, California (1994) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Hours/ | Relative Activity by Month | | | | | | | | | | | |
| Operation | Acre | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
| CULTURE: | | | | | | | | | | | | | |
| Plow Field | 0.20 | | | | | | | | | | 1 | | |
| Subsoil | 0.20 | | | | | | | | | | 1 | | |
| Land Plane Field | 0.34 | | | | | | | | | | 1 | | |
| List Beds | 0.28 | | | | | | | | | | 1 | | |
| Cultivate Beds - 2X | 0.37 | | | 1 | | 1 | | | | | | | |
| Plant Corn & Apply Fertilizer | 0.33 | | | | 1 | | | | | | | | |
| Break Crust | 0.02 | | | | 1 | | | | | | | | |
| Open Ditch - 2X | 0.10 | | | | | 1 | | | | | | | |
| Close Ditch - 2X | 0.10 | | | | | 1 | | | 1 | | | | |
| Apply Insecticide | 0.03 | | | | | 1 | | | | | | | |
| Furrow Out & Sidedress Fertilizer | 0.29 | | | | | 1 | | | | | | | |
| Apply Herbicide | 0.08 | | | | | 1 | | | | | | | |
| Apply Miticide | 0.16 | | | | | | 1 | | | | | | |
| HARVEST: | | | | | | | | | | | | | |
| Combine Corn | 0.22 | | | | | | | | | 1 | | | |
| Bankout Grain | 0.22 | | | | | | | | | 1 | | | |
| Chop Stubble | 0.22 | | | | | | | | | 1 | | | |
| POSTHARVEST: | | | | | | | | | | | | | |
| Disc Stubble | 0.22 | | | | | | | | | 1 | | | |

*Table 6.2 Crop Categories and Data Sources*

| Category | Data Source |
|---|---|
| Alfalfa | Sample Costs to Produce Alfalfa Hay in Sacramento Valley (1992) |
| Almonds | Sample Costs to Produce Almonds in Sacramento Valley (1995) |
| Corn | Sample Costs to Produce Field Corn in Yolo County (1994) |
| Melons | Sample Costs to Produce Mixed Melons in San Joaquin Valley (1992) |
| Oat Hay | Sample Costs to Produce Double Cropped Oat Hay in San Joaquin Valley (1990) |
| Pears | Sample Costs to Produce Pears in Lake County (1994) |
| Prunes | Sample Costs to Produce Prunes in Sacramento Valley (1995) |
| Rice | Sample Costs to Produce Rice in Sutter, Yuba, Placer, and Sacramento Counties (1992) |
| Safflower | Sample Costs to Produce Safflower in Yolo County (1996) |
| Sugar Beets | Sample Costs to Produce Sugar Beets in Yolo County (1994) |
| Tomato | Sample Costs to Produce Tomatoes in the San Joaquin Valley (1992) |
| Walnuts | Sample Costs to Produce English Walnuts in Sacramento Valley (1995) |
| Wheat | Sample Costs to Produce Wheat in Yolo County (1995) |
| Wine Grapes | Sample Costs to Produce Wine Grapes in Sonoma County (1992) |

Therefore, when we could not find the sample production cost estimates for a crop in a particular county, we would use the estimates for the crop in a neighboring county or valley. In cases where we could not decide which estimates were appropriate to use, the experts in the U.C. Cooperative Extension at Davis provided guidance. Table 6.2 lists the major crop categories used in this study. The production cost estimates for major crops in the three counties, Sacramento, Solano, and Yolo, are presented in Tables 6.3-6.5. Using the cost estimates and the agriculture land use data, we computed in GIS the number of mechanical hours needed for crops in each model grid cell and the total hours for a county. Dividing the hours for each cell by the county total, we got the fraction for the cell, which is the factor for the spatial and temporal (monthly) allocation.

To investigate the possibility of obtaining data for weekly and diurnal allocation, we searched many sources and contacted experts in the field, including professors in the Agricultural Economics Department at UC Davis, the Agriculture Economics Library at UC Davis, the agriculture cooperative extension, the agriculture commissioners office in Yolo county, the United States Department of Agriculture office in Davis, the farm service agency in Woodland, and the employment development departments in Woodland and Sacramento. We couldn't find the data for the purpose. The data for weekly and diurnal allocation can be collected by a massive survey of farmers, which is beyond the resources available for this project.

*Table 6.3. Production Cost Estimates by Major Crops (Sacramento County)*

| **Sacramento County** | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Major Crops | PRODUCTION COST (HOURS/ACRE) | | | | | | | | | | | | |
| | Year | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
| Grapes | 5.080 | 0.000 | 0.650 | 0.448 | 0.448 | 0.948 | 1.488 | 0.948 | 0.000 | 0.000 | 0.000 | 0.000 | 0.150 |
| Pears | 10.460 | 0.000 | 0.500 | 0.990 | 2.715 | 2.215 | 1.315 | 1.565 | 1.160 | 0.000 | 0.000 | 0.000 | 0.000 |
| Tomatoes | 4.070 | 0.290 | 0.670 | 0.390 | 0.000 | 0.000 | 0.230 | 1.190 | 0.200 | 0.000 | 0.000 | 1.100 | 0.000 |
| Sudan Seed | 2.970 | 0.000 | 0.000 | 0.000 | 0.460 | 0.300 | 0.000 | 0.000 | 0.815 | 0.815 | 0.580 | 0.000 | 0.000 |
| Sugar Beets | 5.490 | 0.460 | 0.000 | 1.000 | 1.200 | 0.450 | 0.000 | 0.000 | 0.000 | 0.660 | 1.320 | 0.000 | 0.400 |
| Corn | 3.410 | 0.000 | 0.000 | 0.185 | 0.350 | 0.735 | 0.160 | 0.000 | 0.050 | 0.910 | 1.020 | 0.000 | 0.000 |
| Alfalfa | 4.212 | 0.000 | 0.000 | 0.000 | 0.422 | 0.422 | 0.422 | 0.712 | 1.672 | 0.562 | 0.000 | 0.000 | 0.000 |
| Rice | 2.500 | 0.000 | 0.000 | 0.000 | 1.000 | 0.140 | 0.000 | 0.000 | 0.000 | 0.000 | 1.360 | 0.000 | 0.000 |
| Safflower | 2.330 | 0.000 | 0.720 | 0.530 | 0.000 | 0.000 | 0.000 | 0.000 | 0.500 | 0.470 | 0.110 | 0.000 | 0.000 |
| Wheat | 1.440 | 0.000 | 0.060 | 0.000 | 0.000 | 0.000 | 0.570 | 0.000 | 0.240 | 0.000 | 0.440 | 0.130 | 0.000 |

***Table 6.4 Production Cost Estimates by Major Crops (Solano County)***

| **Solano County** | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Major Crops | PRODUCTION COST (HOURS/ACRE) | | | | | | | | | | | | |
| | Year | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
| Almonds | 6.798 | 0.580 | 0.628 | 0.700 | 0.548 | 0.596 | 0.298 | 0.700 | 1.298 | 1.000 | 0.000 | 0.450 | 0.000 |
| Grapes | 5.080 | 0.000 | 0.650 | 0.448 | 0.448 | 0.948 | 1.488 | 0.948 | 0.000 | 0.000 | 0.000 | 0.000 | 0.150 |
| Pears | 10.460 | 0.000 | 0.500 | 0.990 | 2.715 | 2.215 | 1.315 | 1.565 | 1.160 | 0.000 | 0.000 | 0.000 | 0.000 |
| Prunes | 2.910 | 0.200 | 0.000 | 0.730 | 0.500 | 0.440 | 0.300 | 0.300 | 0.300 | 0.000 | 0.140 | 0.000 | 0.000 |
| Walnuts | 8.770 | 0.000 | 2.770 | 0.000 | 0.800 | 0.800 | 0.550 | 0.550 | 1.050 | 0.000 | 2.000 | 0.250 | 0.000 |
| Corn | 3.410 | 0.000 | 0.000 | 0.185 | 0.350 | 0.735 | 0.160 | 0.000 | 0.050 | 0.910 | 1.020 | 0.000 | 0.000 |
| Alfalfa | 4.212 | 0.000 | 0.000 | 0.000 | 0.422 | 0.422 | 0.422 | 0.712 | 1.672 | 0.562 | 0.000 | 0.000 | 0.000 |
| Safflower | 2.330 | 0.000 | 0.720 | 0.530 | 0.000 | 0.000 | 0.000 | 0.000 | 0.500 | 0.470 | 0.110 | 0.000 | 0.000 |
| Sugar Beets | 5.490 | 0.460 | 0.000 | 1.000 | 1.200 | 0.450 | 0.000 | 0.000 | 0.000 | 0.660 | 1.320 | 0.000 | 0.400 |
| Wheat | 1.440 | 0.000 | 0.060 | 0.000 | 0.000 | 0.000 | 0.570 | 0.000 | 0.240 | 0.000 | 0.440 | 0.130 | 0.000 |
| Tomatoes | 4.070 | 0.290 | 0.670 | 0.390 | 0.000 | 0.000 | 0.230 | 1.190 | 0.200 | 0.000 | 0.000 | 1.100 | 0.000 |

***Table 6.5 Production Cost Estimates by Major Crops (Yolo County)***

| Yolo County | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Major Crops | PRODUCTION COST (HOURS/ACRE) | | | | | | | | | | | | |
| | Year | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
| Almonds | 6.798 | 0.580 | 0.628 | 0.700 | 0.548 | 0.596 | 0.298 | 0.700 | 1.298 | 1.000 | 0.000 | 0.450 | 0.000 |
| Grapes | 5.080 | 0.000 | 0.650 | 0.448 | 0.448 | 0.948 | 1.488 | 0.948 | 0.000 | 0.000 | 0.000 | 0.000 | 0.150 |
| Prunes | 2.910 | 0.200 | 0.000 | 0.730 | 0.500 | 0.440 | 0.300 | 0.300 | 0.300 | 0.000 | 0.140 | 0.000 | 0.000 |
| Walnuts | 8.770 | 0.000 | 2.770 | 0.000 | 0.800 | 0.800 | 0.550 | 0.550 | 1.050 | 0.000 | 2.000 | 0.250 | 0.000 |
| Melons | 2.790 | 0.000 | 0.695 | 0.835 | 0.000 | 0.970 | 0.000 | 0.000 | 0.290 | 0.000 | 0.000 | 0.000 | 0.000 |
| Corn | 3.410 | 0.000 | 0.000 | 0.185 | 0.350 | 0.735 | 0.160 | 0.000 | 0.050 | 0.910 | 1.020 | 0.000 | 0.000 |
| Alfalfa | 4.212 | 0.000 | 0.000 | 0.000 | 0.422 | 0.422 | 0.422 | 0.712 | 1.672 | 0.562 | 0.000 | 0.000 | 0.000 |
| Oats | 2.170 | 0.000 | 0.000 | 0.010 | 0.000 | 0.000 | 1.210 | 0.000 | 0.000 | 0.000 | 0.000 | 0.550 | 0.400 |
| Rice | 2.500 | 0.000 | 0.000 | 0.000 | 1.000 | 0.140 | 0.000 | 0.000 | 0.000 | 0.000 | 1.360 | 0.000 | 0.000 |
| Safflower | 2.330 | 0.000 | 0.720 | 0.530 | 0.000 | 0.000 | 0.000 | 0.000 | 0.500 | 0.470 | 0.110 | 0.000 | 0.000 |
| Sugar Beets | 5.490 | 0.460 | 0.000 | 1.000 | 1.200 | 0.450 | 0.000 | 0.000 | 0.000 | 0.660 | 1.320 | 0.000 | 0.400 |
| Wheat | 1.440 | 0.000 | 0.060 | 0.000 | 0.000 | 0.000 | 0.570 | 0.000 | 0.240 | 0.000 | 0.440 | 0.130 | 0.000 |
| Tomatoes | 4.070 | 0.290 | 0.670 | 0.390 | 0.000 | 0.000 | 0.230 | 1.190 | 0.200 | 0.000 | 0.000 | 1.100 | 0.000 |

# 7. Construction Mobile Equipment

## Overview

The original source categories are light-duty industrial mobile equipment and heavy-duty industrial mobile equipment. These categories are used to inventory the combustion emissions from off-road industrial equipment. This study concentrates on non-road mobile equipment used in the construction industries. Examples of such equipment include backhoes, forklifts, loaders, cranes, crawlers, and excavators.

To spatially allocate the emissions estimates for the category, one needs to know the spatial distribution of the activities, that is, where the equipment is used. Though information on the activities can be obtained from surveying the users of the equipment, there are several disadvantages with the survey approach. First, surveys are costly, and the response rate may be low. Second, the equipment is 'mobile,' which means that it is usually used in different places at different times. It is unlikely that all users keep detailed records of where and when the equipment was used. Third and most importantly, the ARB requires that, whenever possible, the allocation surrogate should be based on data that are widely available and can be collected and/or updated routinely. Section 7.2 discusses the spatial surrogate and methods of spatial allocation for this category. Section 7.3 discusses temporal allocation which is based on data from user surveys.

## Spatial Surrogate for Allocation

### The Spatial Surrogate

In this study a spatial allocation surrogate is an indicator measuring the level of emission source activities. The selection of a surrogate is largely based on data availability. One data source that provides information on where a construction project takes place is the building permit records maintained by a county or city building department. The record contains information such as the location of a construction site, valuation of the project, building codes, and date permit issued. Usually the record includes the Assessor's Parcel Number (APN) as well. The permit records are a good data source for spatially allocating construction equipment emissions. The site address or APN can be used to geocode the location of a construction site. The valuation can be used to measure the construction activity level. The issuing date is useful for selecting records for a particular study period. Moreover, the building permit records are public information which are easily available and updated on a regular basis. Therefore, the valuation of construction project per grid cell is selected as the surrogate for spatial allocation of construction mobile equipment emissions.

### Spatial Allocation Procedure

The allocation procedure has three steps. The first step is to collect building permit data and assessor's maps from the county and city building departments in the study area. If the maps are not in digital form, digitize them and index the polygons using APN, then match building permit records to the polygons on the map by APN. Compute the sum of valuations of construction projects for each polygon. The second step is to disaggregate the study area into small zones. Conventional square grids (4km by 4km) are used as the spatial unit for allocation and modeling. Data layers are integrated with the model grid coverage, and then grid cell-based data are computed. In the third step, the allocation factor (weight) for each cell is computed using the valuation. The spatial allocation surrogate can be estimated using regression models, in which the dependent variable is the project valuation per cell and the independent variables are from data on land use, highways, population and employment densities.

### Building Permit Data

The building permit data are maintained by county and city building departments. In general, the county provides building permit service for unincorporated areas of the county. The cities operate their own building departments and issue their own permits. The valuation of a construction project is determined by a trained and certified Chief Building Official (CBO) for the purpose of fairly assessing permit fees and other fees. Valuations shown on the reports are not necessarily the market values of the projects and may vary from jurisdiction to jurisdiction. The form and availability of information may also vary from jurisdiction to jurisdiction. For example, the Yolo County Building Department has all the information we requested, while the City of Woodland' Building Department doesn't include APNs in their records.

The number of building permit records in a county for a given year can be very large. For example, the Sacramento County Building Department issued more than 20,000 building permits in 1996. Due to the large number of records, it is too time-consuming to geocode locations of all construction sites by address matching. Thus, we chose to use the assessor's map and the APN to do spatial matches. The APN serves as a spatial indicator which refers the location of a property to the assessor's map. An APN consists of three parts: the first part (called book index) refers to the map book; the second part refers to the page in the map book; and the third refers to a particular parcel on that page. For example, suppose that the address of a construction site is 550 Jefferson, West Sacramento and the corresponding APN number is 010 5490 120. This APN refers the site to map book 10, page 5490, and parcel 120. The example illustrates the hierarchical structure of the assessor's map: the map book level, the page level, and the parcel level. Using the APN, the location of a construction site can be easily found on the assessor's map. In this study we geocoded the locations of the construction sites to the first level of the assessor's maps. Figure 7.1 shows the assessor's map for Sacramento county. Figure 7.2 is the map for Yolo county. Both maps show boundaries of zones at the map book level. The numbers printed on the polygons are the book indices.

Information Systems (1997), a company located in Grass Valley, California, maintains a database of building permit reports and collects such data from building departments in California on a regular basis. The company provided us data from building permits issued in 1996 from Sacramento county and most of Yolo county but had no data for Solano county at the present. The company didn't have the data for three of the cities in Yolo county, namely the City of Davis, City of Woodland, and City of Winters.
The data came in EXCEL format and contained the information shown in Table 7.1.

In the next step we collected building permit data from the three cities in Yolo county. As mentioned previously, the City of Woodland Building Department could provide us building permit records issued in 1996 as requested but no Assessor's parcel numbers on the records. Using a street map of Woodland provided by the building inspector on which the boundaries of zones at the book index level were marked, we assigned APN to the construction projects with valuation larger than $5000, and aggregated the valuations to the zones at the book index level. The City of Davis provided building permit reports of all new buildings constructed in 1996. Although there were no valuations for the projects, we were given a formula to estimate the valuation of a project based on the building square footage which was available. The building permits issued in 1996 by the City of Winters also lacked information on valuations of the projects. We estimated the valuations based on building permit surcharges using a formula given by the city. Due to time constraints we were unable to collect building permit data from Solano county and the cities in that county.

***Table 7.1 Building Permit Data***

| FIELD NAME | DESCRIPTION |
|---|---|
| ApplyDate | Date of permit Application |
| IssueDate | Date of Permit Issuance |
| Permit | Permit Number assigned by building departments |
| APN | Assessor's parcel number |
| Site | Address of the construction project |
| Type | Type of Construction |
| Value | Valuation assigned by building official |
| SF | Square footage of project |
| BuildDept | Building department name |

*Figure 7.1 Sacramento County Assessor's Map*

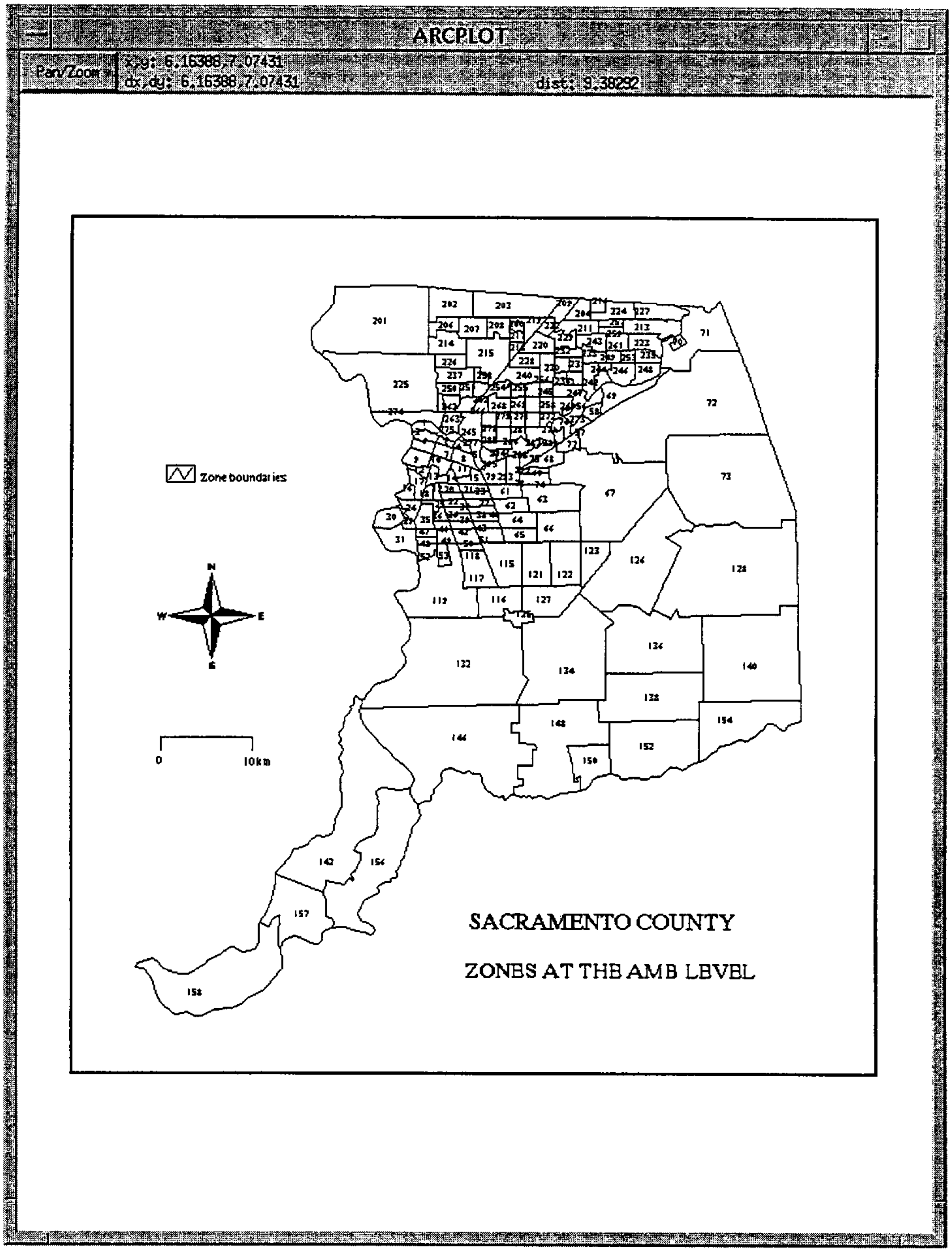
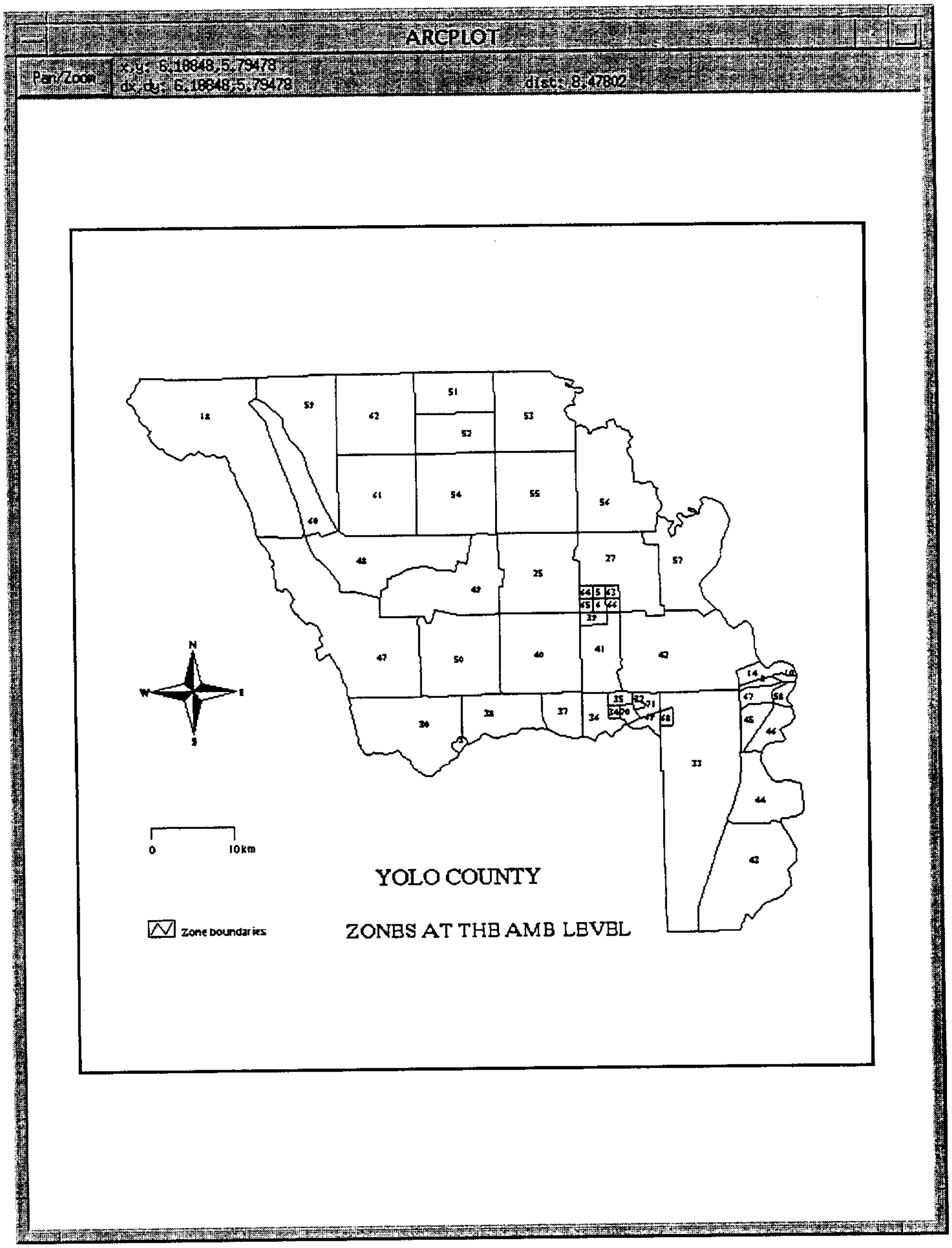


*Figure 7.2 Yolo County Assessor's Map*



### Regression Model Estimation

Regression models are estimated using the least square estimation method. The dependent variable is the valuation per 4km plot. The predictors are the attributes of the plot, such as percentage of urban land use, miles of highways, population density, and employment density. The models are estimated using the data from two counties, Sacramento county and Yolo county. The variables used in the models are defined in Table 7.2 below.

***Table 7.2 Definition of Variables***

| Variable Name | Definition |
|---|---|
| Dependent variable | |
| VALUE | Project valuation in $1000 |
| Predictors | |
| URBAN | Percentage of urban land use |
| HWY | Miles of highway |
| POP | Population density (1000 persons/square miles) |
| CSTEMP | Construction employment density (1000 employees/square miles) |

Model 1, shown in Table 7.3, has only one predictor - construction employment density. The adjusted R square is 0.2258, indicating that with a single predictor about 22 percent of the variations in the data is accounted for by the model. The model shown in Table 7.4 has 3 predictors - URBAN, HWY, and POP. The adjusted R square is increased from 0.2258 to 0.3867, and the Root MSE (mean square error) decreases from 8787.4 to 7820.9. Thus, the model in Table 7.4 fits the data much better than does the model in Table 7.1. In the second model the coefficients for URBAN and HWY have positive signs and the coefficient for POP has a negative sign. It indicates that construction activities are more likely to occur in these parts of the urban areas where highway accessibility is good and population density is relatively lower. In other words, it is expected that the value of the surrogate for spatially allocating construction mobile equipment emission would be higher in new development areas which are usually at the edge of the city. Besides the two models discussed above, we estimated another model with all four variables CSTEMP, URBAN, HWY, and POP. It turns out that the coefficient of CSTEMP is not statistically significantly different from zero ($t = 0.57$) and that the adjusted R square decreases to 0.3857. Thus, it is suggested that CSTEMP not be included in the selected model. A comparison between the average predicted and observed values is presented in Figure 7.3. The averages are computed based on six prediction intervals: (0.0 to 4.9), (5.0 to 9.9), (10.0 to 14.9), (15.0 to 19.9), (20.0 to 24.9), and (25.0 or larger). The values displayed are in $1000. The plot shows that the difference between the average predicted and observed values are reasonably small.

***Table 7.3 Estimation Results - Model 7.1***

| | |
|---|---|
| Regression Model | Number of obs = 390 |
| | F( 1, 388) = 114.44 |
| | Prob > F = 0.000 |
| | R-square = 0.2278 |
| | Adj R-square = 0.2258 |
| | Root MSE = 8787.4 |

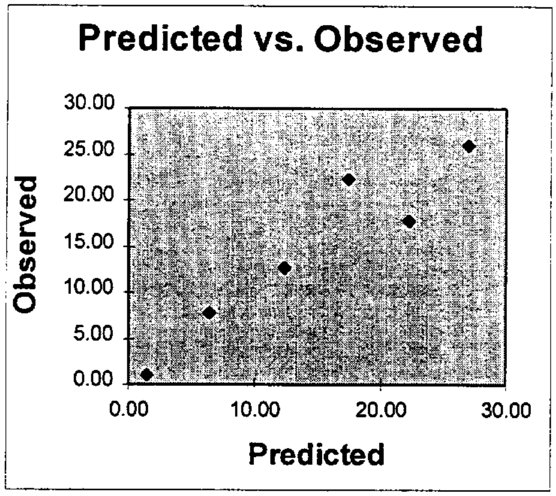
| Variable | Coefficient | Std. Err | Asy. T ratio | P\|>t\| |
|---|---|---|---|---|
| CSTEMP | 105160.9 | 9830.1 | 10.698 | 0.000 |
| CONS | 2278.1 | 477.6 | 4.769 | 0.000 |

***Table 7.4 Estimation Results - Model 7.2***

| | |
|---|---|
| Regression Model | Number of obs = 390 |
| | F( 3,388) = 82.77 |
| | Prob > F = 0.000 |
| | Rsquare = 0.3915 |
| | Adj R-square = 0.3867 |
| | Root MSE = 7820.9 |

| Variable | Coefficient | Std. Err | Asy. T ratio | P\|>t\| |
|---|---|---|---|---|
| URBAN | 401.89 | 53.09 | 7.56 | 0.000 |
| HWY | 899.22 | 206.85 | 4.34 | 0.000 |
| POP | -3652.97 | 991.50 | -3.68 | 0.008 |
| CONS | 283.06 | 524.13 | 0.54 | 0.589 |

*Figure 7.3 Predicted vs. Actual Valuations (Construction Mobile Equipment)*




## Temporal Activity Profiles and Ambient Temperature Effects

### The Survey

The data for temporal allocations were obtained by surveying a sample of companies that use construction mobile equipment. From the Business Prospector we obtained a list of local companies in the construction industries by the SIC codes shown in Table 7.5. There are 880 construction companies on the list. In addition, we searched Pacific Bell Yellow Pages and contacted the Sacramento Builders Exchange - a local trade organization for the construction business. The final list consisted of 1000 or so companies, including home and building contractors, road, bridge, and trench contractors, excavators, and other miscellaneous contractors.

*Table 7.5 SIC Codes for the Construction Industries*

| SIC Codes | Description |
|---|---|
| 1521 | Single family housing construction |
| 1542 | Non-residential construction |
| 1611 | Highway and street construction |
| 1622 | Bridge, tunnel, & elevated highway |
| 1623 | Water, sewer, and utility lines |
| 1629 | Heavy construction |

The telephone survey was conducted during May and June of 1997. The questionnaire used for the survey was carefully prepared. Reference materials, telephone conversations with the equipment users, and the questionnaires used for previous categories all helped prepare the survey questions. The survey questions focused on temporal variations of the equipment use, and the effects of temperature and weather on the usage. A copy of the questionnaire is included in the Appendix.

Before selecting a sample, we did a test survey. The list from the Business Prospector included information on company size in terms of number of employees. In the test survey we found that the smaller construction companies (less than 3 employees) usually did not have construction mobile equipment. Therefore, we excluded those tiny companies from the list. For Solano county and Yolo county we selected all construction companies with more than 3 employees. Due to the large number of construction companies in Sacramento county, we selected all companies with more than 15 employees plus a random sample of companies with 4 to 15 employees. The companies selected above, plus the ones from the Sacramento Builders Exchange list and some found in the phone books, made up a survey sample of about 200 companies. We surveyed the 200 or so companies and were able to obtain 60 completed questionnaires. The spatial distribution of the respondents is 34 in Sacramento, 17 in Solano, and 9 in Yolo.

### Temporal Activity Profiles and Temperature Effects

The temporal activity profiles and temporal allocations are summarized in Tables 7.6-7.9. The survey results show that the seasonal variation in the use of non-road construction mobile equipment is significant. As expected, there are much less construction activities in the winter season than the other seasons. The seasonal allocation is 13.4 % for the winter, 29.9% for the Spring, 31.1% for the Summer, and 25.6% for the Fall. The weekly pattern (day of the week) is the equipment is used most frequently during weekdays (91.9%). The weekly allocation is 18.4% for each weekday (Monday to Friday), 5.7% for Saturday, and 2.3% for Sunday. The diurnal pattern is 55.3% of the activity takes place in the morning (6 a.m. to noon), 43.6% in the afternoon (noon to 6 p.m.), and 1.1% in the evening (after 6 p.m.).

The monthly allocation given in Table 7.7 was made based on the seasonal estimates, assuming that the source activity is uniformly distributed within the months of a season. Each season is assumed to have three months, specifically winter (December, January, February), spring (March, April, May), summer (June, July, August), and fall (September, October, November). Weekly allocations were computed under the assumption that the source activity during weekdays is homogenous. The hourly allocations were calculated based on the work schedules reported by the respondents.

The effects of ambient temperature and weather are shown in Table 7.10. The survey results indicate that temperature and weather do have impact on the use of non-road mobile construction equipment. This is expected because construction mobile equipment is mostly used outdoors. 41.7% of the respondents reported that they did alter their work schedule on hot days during the summer. On those days, they would typically start work earlier in the morning and end the work day earlier in the afternoon. Most of them would shift their schedule one hour earlier, from the normal schedule of 7:00 a.m. to 3:30 p.m. to 6:00 a.m. to 2:30 p.m. As expected, not all construction workers work on rainy days. The survey shows that, while 67% worked on rainy days, only 55% actually used the equipment on those days. Moreover, on rainy days, the use of the equipment was only a fraction of that of non-rainy days (23.3% on average).

### Confidence Intervals

The confidence intervals for the estimates on quarterly, weekly, and diurnal allocations are presented in Tables 7.11-7.13. The confidence intervals for estimates of temperature and weather effects estimates are given in Table 7.14.

*Table 7.6 Temporal Activity Profiles (Construction Mobile Equipment)*

| CATEGORY | QUARTERLY OR SEASONAL | | | | WEEKLY | | | DIURNAL | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Winter | Spring | Summer | Fall | Monday to Friday | Saturday | Sunday | 6 a.m.- 12 p.m. | 12p.m.- 6p.m. | 6 p.m. - 0 a.m. |
| Construction Equipment | 13.4 | 29.9 | 31.1 | 25.6 | 91.9 | 5.7 | 2.4 | 55.3 | 43.6 | 1.1 |

*Table 7.7 Monthly Activity Profiles (Construction Mobile Equipment)*

| CATEGORY | MONTHLY ACTIVITY PROFILE (%) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
| Construction Equipment | 4.5 | 4.5 | 10.0 | 10.0 | 10.0 | 10.4 | 10.4 | 10.4 | 8.5 | 8.5 | 8.5 | 4.5 |

*Table 7.8 Weekly Activity Profiles (Construction Mobile Equipment)*

| CATEGORY | WEEKLY ACTIVITY PROFILE (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday |
| Construction Equipment | 18.4 | 18.4 | 18.4 | 18.4 | 18.4 | 5.7 | 2.4 |

*Table 7.9 Diurnal Activity Profiles (Construction Mobile Equipment)*

| CATEGORY | HOURLY ACTIVITY PROFILE (%) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0-1 | 1-2 | 2-3 | 3-4 | 4-5 | 5-6 | 6-7 | 7-8 | 8-9 | 9-10 | 10-11 | 11-12 |
| Construction Equipment: Normal Days | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.0 | 10.5 | 10.7 | 10.7 | 10.7 | 10.7 |
| Construction Equipment Hot Days | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.8 | 6.0 | 10.8 | 10.9 | 10.9 | 10.9 | 10.9 |
| | 12-13 | 13-14 | 14-15 | 15-16 | 16-17 | 17-18 | 18-19 | 19-20 | 20-21 | 21-22 | 22-23 | 23-24 |
| Construction Equipment: Normal Days | 10.7 | 10.7 | 10.6 | 7.0 | 3.1 | 1.5 | 0.5 | 0.4 | 0.2 | 0.0 | 0.0 | 0.0 |
| Construction Equipment Hot Days | 10.9 | 10.3 | 8.2 | 4.6 | 2.2 | 1.1 | 0.6 | 0.4 | 0.2 | 0.0 | 0.0 | 0.0 |

***Table 7.10 Temperature and Weather Effects (Construction Mobile Equipment)***

| CATEGORY | TEMPERATURE EFFECTS | | | WEATHER EFFECTS | | |
|---|---|---|---|---|---|---|
| | Alter Work Schedule on Hot Days | Start Work Earlier in the Morning | Work in the Evening | Work on Rainy Days | Use Equipment on Rainy Days | Average Usage as Fraction of Non-rainy Day |
| Construction Equipment | 41.7 | 41.7 | 0.0 | 67.0 | 55.0 | 23.3 |

*Table 7.11 Confidence intervals For Seasonal Estimates (Construction Mobile Equipment)*

| Period | Mean | 95% Confidence Interval |
|---|---|---|
| Winter | 13.4 | [ 11.1, 15.7 ] |
| Spring | 29.9 | [ 28.4, 31.4 ] |
| Summer | 31.1 | [ 29.5, 32.7 ] |
| Fall | 25.6 | [ 24.2, 27.0 ] |

Notes: n = 58, value = percentages

*Table 7.12 Confidence intervals For Weekly Estimates (Construction Mobile Equipment)*

| Period | Mean | 95% Confidence Interval |
|---|---|---|
| Monday - Friday | 20.0 | N/A |
| Saturday | 1.25 | [ 0.92, 1.58 ] |
| Sunday | 0.52 | [ 0.25, 0.79 ] |

Notes: n = 60, value = days per month

*Table 7.13 Confidence intervals For Diurnal Estimates (Construction Mobile Equipment)*

| Period | Mean | 95% Confidence Interval |
|---|---|---|
| Morning | 5.2 | [ 5.1, 5.3 ] |
| Afternoon | 4.1 | [ 3.9, 4.3 ] |
| Evening | 0.1 | [ 0.0, 0.2 ] |

Notes: n = 59, value = hours per day

*Table 7.14 Confidence intervals For Temperature and Weather Effects (Construction Mobile Equipment)*

| | Altering Procedures on Hot Days | Altering Procedures on Rainy Days |
|---|---|---|
| Count | 25 | 33 |
| Relative Frequency | 0.416 | 0.550 |
| Standard Deviation | 0.063 | 0.064 |
| 95% Confidence Intervals | [ 0.293, 0.539 ] | [ 0.425, 0.675 ] |

Notes: n = 60

# 8. Trains

## Overview

The trains categories are used by the ARB to inventory the emissions from the combustion of diesel fuel by trains during road haul and switching operations. Booz-Allen & Hamilton Inc. has done a detailed estimate of locomotive emissions in California for the ARB. The objective of this study is to come up with a simple and low-cost method to allocate county-level train emission estimates to model grid cells.

## Spatial Surrogate for Allocation

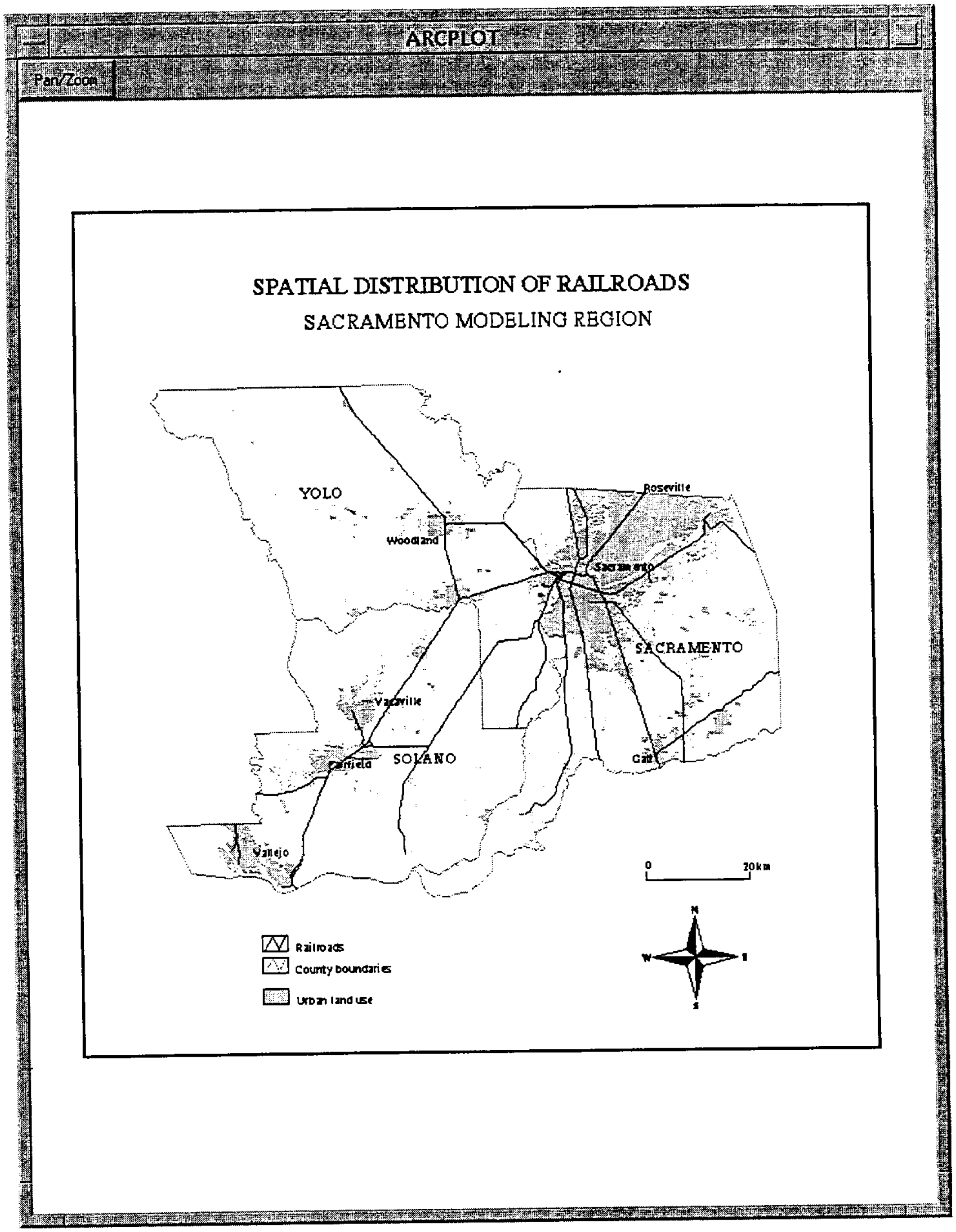
Train emissions result from railroad activities. Railroad activity levels in a region are determined by many factors such as total number of trains operated, intensity of local and yard operations, the average HP and trailing tons of each train, and the geography and terrain of the region. In this study we select a simple spatial surrogate for allocation. We use miles of railways to measure railroad activity levels and to compute allocation factors for each model grid cell accordingly. Specifically, the allocation factor is obtained by first measuring the length of railways in a model grid cell and then dividing the length by the total length of railways in the study area. The information on spatial distribution of railways is available from the TIGER files. Since the TIGER files are easily available and updated on a regular basis by the U.S. Census Bureau, it is not necessary for us to estimate a model to predict the distribution of railroads based on other data.

The railway coverage used for computing the allocation factors was extracted from the street coverage which had been created from the 94 TIGER/Line files. The procedure of converting TIGER files to Arc/Info coverages is described in ESRI (1996). To extract the railway lines from the street coverage was a straightforward operation. In ArcEdit the railway lines were first selected, then the "put" command was used to create a new coverage using those lines. To compute the allocation factors, the railway coverage was overlaid with the model grid coverage (a line in polygon operation). An AML program was written to calculate the length and fraction of length of railways in each grid cell. The fraction is the factor for allocation. The map of spatial distribution of the railways in the study area is presented in Figure 8.1.

## Temporal Activity Profiles and Ambient Temperature Effects

To obtain data for temporal allocation, it's necessary to survey the railway companies that have rail operations in the region. However, we were unable to obtain the cooperation of the railway companies. The temporal activity profiles could not be developed due to the difficulty in data collection.

*Figure 8.1 Map of Spatial Distribution of Railroads in the Study Area*



## 9. Summary and Conclusions

Information on spatial and temporal distributions of emissions is essential for developing emission inventories and ozone air quality simulation models such as the Sacramento State Implementation Plan (SIP) Urban Airshed Model (UAM). The objective of this project is to identify and investigate important temporal and spatial variations in emissions in the Sacramento modeling region, specifically from non-road mobile sources and industrial surface coatings and related process solvents. The emission categories included in this study are shown in Table S.1. The study region includes three counties, namely Sacramento County, Solano County, and Yolo County, in the State of California.

***Table S.1 Emission Source Categories Included in This Study***

| EMISSION TYPE | CATEGORY |
|---|---|
| Industrial Surface Coatings and Related Process Solvents | Auto Refinishing |
| | Adhesives and Sealants |
| | Can & Coil Coatings / Metal Parts and Products Coatings |
| Non-Road Mobile | Farm Equipment |
| | Construction Equipment |
| | Train |

To accomplish the objective, five specific tasks were performed along the following three work lines:

- development of spatial allocation surrogates
- development of temporal activity profiles
- development of estimates of ambient temperature effects

## Development of Spatial Allocation Surrogates

The purpose was to develop methods and activity indicators to spatially allocate countywide emission estimates to model grid cells. As required by the ARB, the spatial allocation surrogates must be based on parameters or data that are easily available and can be updated on a regular basis. The new approach developed in this study consisted of selecting a spatial activity indicator for a given source category, collecting data, using a Geographical Information System (GIS) to evaluate the spatial distribution of the indicator and compute allocation factors for a chosen spatial unit (4km by 4km model grid cell). Notice that this study uses the activity indicators as the spatial allocation surrogates. Table S.2 shows the activity indicators selected for the source categories included in this study and the methods for estimating and updating them.

The allocation surrogates selected are fairly simple, such as the number of auto refinishing facilities, or miles of railroads, or hours of farm equipment used per spatial unit. Information on spatial distribution of railroads is available from U.S. Census Bureau's topographically integrated geographic encoding and referencing (TIGER) files and is updated by the bureau regularly. Data on farm equipment usage can be found in the sample production cost estimate reports prepared by the county farm advisors and the University of California Cooperative Extension. The values of spatial surrogates of the remaining categories can be estimated using widely available data such as land uses (available from California Department of Water Resources), population and employment (from population census), and highways (from TIGER files). The equations for estimation or update are presented in Table S.3. The regression equations were estimated using data collected from the study area. Detailed descriptions on data and model estimation are given in the following sections of this report.

## Development of Temporal Activity Profiles

The objective was to develop monthly, weekly (day of the week), and diurnal activity profiles for each source category shown in Table S.1. Since emissions are estimated by multiplying emission factors with activity level, the activity profiles developed can be used to scale annual emission estimates to determine monthly, weekly, or hourly emissions. Except the farm equipment category where monthly activity profile could be estimated from the sample production cost estimates, the temporal activity profiles of the source categories were developed based on data we obtained from the user surveys (we were not able to get data for the train category). Monthly activity profiles are presented in Table S.4 and S.5. Weekly activity profiles are shown in Table S.6. Table S.7 presents diurnal activity profiles. The monthly profiles contain the fraction (in percentage) of annual emissions allocated to each month. The weekly profiles contain the fraction of weekly emissions allocated to each day of the week. The hourly profiles include fraction of daily emissions allocated to each hour of the day. The data collection processes and assumptions made in developing the profiles are discussed in the sections for the individual categories. Confidence intervals for the estimates are also reported in the sections.

**Development of Estimates of Ambient Temperature Effects**

The effects of high temperature (90 °F or above) and weather (raining or not) on the source activities are summarized in Table S.8. The data came from the surveys. Two effects of temperature and weather can be observed from the table: (1) the direct effect of increasing temperature or rain (e.g., most auto refinishers change paint formulation on hot days and/or rainy days); (2) indirect effect of changes in activity patterns which demonstrate significant time shifts to account for high ambient temperatures (e.g., 41.7% of respondents in the construction industries reported altering work schedules on hot days). Confidence intervals for the estimates are given in the sections for the individual categories.

*Table S.2 Approaches to Estimation of Spatial and Temporal Activity Profiles*

| CATEGORY | ACTIVITY INDICATORS | ESTIMATION/ UPDATE METHOD | PREDICTORS | TEMPORAL PROFILES |
|---|---|---|---|---|
| Auto Refinishing | Number of the facilities per grid square | Poisson regression | % urban land use, highway miles, population density, retail employment density | Quarterly, weekly, diurnal |
| Adhesives and Sealants | Number of the facilities per grid square | Poisson Regression | % urban land use, highway miles, population density, manufacturing employment density | Quarterly, weekly, diurnal |
| Can & Coil Coatings / Metal Parts and Products Coatings | Number of the facilities per grid square | Poisson Regression | % urban land use, highway miles, population density, manufacturing employment density | Quarterly, weekly, diurnal |
| Farm Equipment | Farm equipment use hours/grid square | Updated production cost estimates | | Monthly |
| Construction Equipment | Construction valuation per grid square | Regression | % urban land use, highway miles, population density | Seasonal, weekly, diurnal |
| Train | Miles of railroads per grid square | Updated TIGER files | | |

***Table S.3 Summary of Regression Models for Estimating Spatial Activities***

| CATEGORY | MODEL TYPE | ACTIVITY INDICATOR | PREDICTION EQUATION | $b_0 + \sum_j b_j x_{ij}$ | $R^2$ |
|---|---|---|---|---|---|
| Auto Refinishing | Poisson regression | Number of the facilities per grid square* | $\hat{Y}_i = \exp(b_0 + \sum_j b_j x_{ij})$ | $-2.681 + 0.041x_1 + 0.146x_2 + 0.382x_3 + 3.103x_4 - 1.274x_6$ | 0.60 |
| Adhesives and Sealants | Poisson Regression | Number of the facilities per grid square | $\hat{Y}_i = \exp(b_0 + \sum_j b_j x_{ij})$ | $-2.379 + 0.055x_1 + 0.173x_2 - 0.356x_3 + 9.494x_5 - 1.625x_7$ | 0.51 |
| Can & Coil Coatings / Metal Parts and Products | Poisson Regression | Number of the facilities per grid square | $\hat{Y}_i = \exp(b_0 + \sum_j b_j x_{ij})$ | $-4.111 + 0.034x_1 + 0.087x_2 + 30.128x_5 - 6.491x_7$ | 0.33 |
| Construction Equipment | Regression | Construction valuation ($1000) per grid square | $\hat{Y}_i = b_0 + \sum_j b_j x_{ij}$ | $283.06 + 401.89x_1 + 899.22x_2 - 3652.97x_3$ | 0.39 |

*The modeling unit is a 4 km by 4km grid square.

$x_1$: Percentage of urban land use,
$x_2$: Miles of highways
$x_3$: Population density (1000 persons/square miles)
$x_4$: Retail employment density (1000/sqaure miles)
$x_5$: Manufacturing employment density (1000 /square miles)

$x_6$=x3*x4
$x_7$=x3*x5

*Table S.4 Monthly Activity Profiles*

| CATEGORY | MONTHLY ACTIVITY PROFILE | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
| Auto Refinishing | 8.3 | 8.3 | 8.3 | 8.3 | 8.3 | 8.3 | 8.4 | 8.4 | 8.4 | 8.3 | 8.3 | 8.3 |
| Adhesives and Sealants | 8.3 | 8.3 | 8.3 | 8.3 | 8.3 | 8.3 | 8.3 | 8.3 | 8.3 | 8.4 | 8.4 | 8.4 |
| Can & Coil / Metal Parts and Products Coatings | 8.2 | 8.2 | 8.2 | 8.6 | 8.6 | 8.6 | 8.5 | 8.5 | 8.5 | 8.0 | 8.0 | 8.0 |
| Construction Equipment | 4.5 | 4.5 | 10.0 | 10.0 | 10.0 | 10.4 | 10.4 | 10.4 | 8.5 | 8.5 | 8.5 | 4.5 |

Notes: All figures are in percentages.

***Table S.5 Temporal (Monthly) Activity Profile for Farm Equipment***

| COUNTY | EMISSION ALLOCATION FACTORS (%) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Year | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
| Sacramento | 100.00 | 1.36 | 5.99 | 8.17 | 13.76 | 12.99 | 9.95 | 7.41 | 8.26 | 11.17 | 17.26 | 2.66 | 1.00 |
| Solano | 100.00 | 1.69 | 5.83 | 4.29 | 8.49 | 8.50 | 12.02 | 12.31 | 21.90 | 9.57 | 9.95 | 4.89 | 0.57 |
| Yolo | 100.00 | 2.93 | 11.53 | 7.02 | 5.74 | 5.92 | 11.76 | 11.77 | 14.69 | 6.95 | 11.94 | 9.25 | 0.51 |

Note: The allocation factors are computed based on sample production costs (hours/acre) data estimated by the farm advisors and the University of California Cooperative Extension and the agriculture land use data obtained from the California Department of Water Resources.

*Table S.6 Weekly Activity Profiles*

| CATEGORY | WEEKLY ACTIVITY PROFILE | | | | | | |
|---|---|---|---|---|---|---|---|
| | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday |
| Auto Refinishing | 19.5 | 19.5 | 19.5 | 19.5 | 19.5 | 2.4 | 0.1 |
| Adhesives and Sealants | 18.6 | 18.6 | 18.6 | 18.6 | 18.6 | 5.6 | 1.6 |
| Can & Coil / Metal Parts and Products Coatings | 17.8 | 17.8 | 17.8 | 17.8 | 17.8 | 8.3 | 2.6 |
| Construction Equipment | 18.4 | 18.4 | 18.4 | 18.4 | 18.4 | 5.7 | 2.4 |

Notes: All figures are in percentages.

*Table S.7 Diurnal Activity Profiles*

| CATEGORY | HOURLY ACTIVITY PROFILE | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0-1 | 1-2 | 2-3 | 3-4 | 4-5 | 5-6 | 6-7 | 7-8 | 8-9 | 9-10 | 10-11 | 11-12 |
| Auto Refinishing | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 10.2 | 10.2 | 10.2 | 10.2 | 10.2 |
| Adhesives and Sealants | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 |
| Can & Coil / Metal Parts and Products Coatings | 0.4 | 0.4 | 0.4 | 0.4 | 0.9 | 0.9 | 1.9 | 2.6 | 9.4 | 9.4 | 9.4 | 9.4 |
| Construction Equipment: Normal Days | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.0 | 10.5 | 10.7 | 10.7 | 10.7 | 10.7 |
| Construction Equipment Hot Days | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.8 | 6.0 | 10.8 | 10.9 | 10.9 | 10.9 | 10.9 |
| | 12-13 | 13-14 | 14-15 | 15-16 | 16-17 | 17-18 | 18-19 | 19-20 | 20-21 | 21-22 | 22-23 | 23-24 |
| Auto Refinishing | 7.8 | 7.8 | 7.8 | 7.8 | 7.8 | 7.8 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 |
| Adhesives and Sealants | 8.2 | 8.2 | 8.2 | 8.2 | 8.2 | 8.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |
| Can & Coil / Metal Parts and Products Coatings | 9.4 | 9.4 | 9.2 | 9.0 | 8.8 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 0.9 |
| Construction Equipment: Normal Days | 10.7 | 10.7 | 10.6 | 7.0 | 3.1 | 1.5 | 0.5 | 0.4 | 0.2 | 0.0 | 0.0 | 0.0 |
| Construction Equipment Hot Days | 10.9 | 10.3 | 8.2 | 4.6 | 2.2 | 1.1 | 0.6 | 0.4 | 0.2 | 0.0 | 0.0 | 0.0 |

Notes: All figures are in percentages.

*Table S.8 Summary of Temperature and Weather Effects*

| CATEGORY | TEMPERATURE EFFECTS | | | | | WEATHER EFFECTS | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Alter Procedures on Hot Days | On Hot Days | | | | Alter Procedures on Rainy Days | On Rainy Days | | | |
| | | Likely Use in Morning | Likely Use in Evening | Don't Use | Change Formu-lation | | Likely Use in Morning | Likely Use in Evening | Don't use | Change Formu-lation |
| Auto Refinishing | 84.7 | 27.8 | 1.4 | 6.9 | 76.4 | 62.5 | 1.4 | 5.6 | 0.0 | 56.9 |
| Adhesives and Sealants | 10.3 | 1.5 | 0.0 | 1.5 | 7.4 | 14.7 | 1.5 | 4.4 | 0.0 | 8.8 |
| Can & Coil / Metal Parts and Products Coatings | 18.2 | 4.5 | 4.5 | 0.0 | 13.6 | 40.9 | 0.0 | 0.0 | 18.2 | 18.2 |

| | Alter Work Schedule on Hot Days | Start Work Earlier in the Morning | Work in the Evening | Work on Rainy Days | Use Equipment on Rainy Days | Average Usage as Fraction of Non-rainy Day |
|---|---|---|---|---|---|---|
| Construction Equipment | 41.7 | 41.7 | 0.0 | 67.0 | 55.0 | 23.3 |

Notes: 1. All figures are in percentages. 2. Respondents can choose more than one answers to some of the survey questions. 3. Hot days are defined as the days with temperature above 90 degrees Fahrenheit.

# 10. Recommendations

As with the development of many emissions inventories, this study points out the need for additional data. The sample production cost estimates for crop types show only monthly variations in farm equipment usage. To obtain data on weekly and diurnal patterns of the activity, a special survey of farmers is recommended. The ARB is urged to contact the University of California Cooperative Extension to obtain sample production cost estimates for counties beyond the Sacramento modeling region. It is also suggested that the ARB begin a dialogue with the California Department of Water Resources to ensure that most updated land use data is used in developing the inventories. The effort to persuade the railway companies to cooperate in providing data on train operations should continue.

This study has relied on surveys to obtain data for developing temporal activity profiles. Due to non-responses in the surveys and in some cases the difficulty to identify the survey subjects, the samples may not be truly random. This means that the statistical estimates on temporal patterns may not be as accurate as one might wish. It is suggested that the results be verified and possibly corrected using additional data in further studies. The effects of random grid locations on regression results have been evaluated in this study. It is highly recommended that the effects of changing grid cell size on model results be evaluated in further studies.

The GIS approach requires information on locations of the facilities producing emissions. One way to geocode the location of a facility is to match the address of the facility with those in the TIGER files. One concern is that there may be cases where the available addresses refer to company offices or headquarters as opposed to the actual facility producing emissions. In this study the spatial resolution for modeling is a 4 km by 4 km grid cell, within which the emission producing facilities are counted. Thus the use of office addresses may not be a serious problem if the office and the facilities are reasonably close. However, the error would be too large if the office and the facilities are located far apart. In those cases, the company office might be called to obtain the addresses of its facilities. Most business listings available include the name and address of the company, the phone number, and name of contact person.

The statistical modeling methodology proposed in the study is quite general. It can handle continuous data (e.g., the size of a facility) as well as discrete data (e.g., the presence of a facility). When the measurements on the response variable are continuous, such as the size of a facility measured by number of employees, or the number of gallons of paint used in a year, linear regression models can be used. In the section on construction mobile equipment we showed how linear regression models might be applied when the response variable is measured at the interval or ratio scales.

# 11. References


Bachman, W, W.S. Sarasua and R. Guensler (1996). A GIS Framework for Mobile Source Emissions Modeling. Paper presented at the Transportation Research Board 75th Annual Meeting at Washington, D.C.

Battye, W., G. Viconovic, and A. Williams (1993). Inventory Data Base Analysis for Area Source Solvent Emissions. Report prepared for the U.S. Environmental Protection Agency (EPA 68-D2-0181) by EC/R Incorporated, Durham, North Carolina,.

Birkin, M., G. Clarke, M. Clarke, and A. Wilson (1996). Intelligent GIS: Location Decisions and Strategic Planning. GeoInformation International, Cambridge, UK.

Booth, K. (1990). Industrial Packaging Adhesives. CRC Press, Boca Raton, FL.

Booz, Allen and Hamilton Inc (1989). Locomotive Emission Study. Final reports prepared for the California Air Resources Board.

Business Prospector (1996). Business Prospector. Beaverton, OR.

CARB (1991a). Metal Parts and Products Surface Coating Operations. Prepared by the Compliance Assistance Program, California Air Resources Board.

CARB (1991b). Metal Container, Closure, and Coil Coating Operations. Prepared by the Compliance Assistance Program, California Air Resources Board.

CARB (1995). Emission Inventory Procedural Manual (Volume III): Methods for Assessing Area Source Emissions. Prepared by the California Air Resources Board, Sacramento, CA.

Causley, Marianne (1995). Development of an Improved Inventory of Emissions from Pleasure Craft in California. Prepared for the California Air Resources Board (A132-184) by Systems Applications International, San Rafael, CA.

Cook, J.P. (1970). Construction Sealants and Adhesives. John Wiley & Sons, New York.

Duns Marketing Service (1996). Microcosm. Mountain Lakes, New Jersey.

ESRI (1994). Surface Modeling with TIN. Environmental Systems Research Institute, Redlands, California.

ESRI (1996). Network Analysis. Environmental Systems Research Institute, Redlands, California.

Flowerdew, R., and M. Green (1994). Areal Interpolation and Types of Data. In S. Fotheringham and P. Rogerson (eds.), Spatial Analysis and GIS, Taylor & Francis Inc, Bristol, PA.

Fotheringham, S. and P. Rogerson (1994) (eds). Spatial Analysis and GIS. Taylor & Francis Ltd, London, UK.

Fowler, F.J. (1984). Survey Research Methods. Sage Publications, Beverly Hills, CA.

Frey, J.H. (1983). Survey Research by Telephone. Sage Publications, Beverly Hills, CA.

Goodchild, M.F., N. S-N Lam (1980). Real Interpolation: A Variant of the Traditional Spatial Problem. Geo-Processing, 1, 297-312.

Goodchild, M.F., B.O. Parks, and L.T. Steyaert (1993). Environmental Modeling with GIS. Oxford University Press, New York.

Goodchild, M.F., L.T. Steyaert, B.O. Parks, C. Johnston, D. Maidment, M. Crane, and S. Glendinning (1996). GIS and Environmental Modeling: Progress and Research Issues. GIS World Books, Fort Collins, CO.

Green, M. (1990). Statistical Methods for Areal Interpolation, In J. Harts, H. F. L. Ottens and H. J. Scholten (eds.), EGIS '90: Proceedings of the First European Conference on Geographic Information Systems, EGIS Foundation, Utrecht, The Netherlands, 1, pp. 392-399.

Greene, W.H. (1990). Econometric Analysis. Macmillan Publishing Company, New York.

Hogg, R.V. and E.A. Tanis (1988). Probability and Statistical Inference. Macmillan Publishing Company, New York.

Information Systems (1997). Building Permit Reports. Provided by Information Systems, Grass Valley, CA 95945.

Johnson, R.A. and D.W. Wichern (1992). Applied Multivariate Statistical Analysis. Prentice Hall, Englewood Cliffs, New Jersey.

Kakabadse, G. (1984) (eds). Solvent Problems in Industry. Elsevier Applied Science Publishers, New York.

KVB, Inc. (1980). Inventory of Emissions from Non-Automotive Vehicle Sources. Contract No. ARB A6-167-30. Prepared for the California ARB.

Laurini, R. and D. Thompson (1992). Fundamentals of Spatial Information Systems. Academic Press, New York.

Livingston, P. (1996). Personal communications with Pete Livingston, Staff Research Associate, Department of Agriculture Economics, Cooperative Extension, University of California, Davis, CA 95616.

Longley, P. and M. Batty (1996) (eds). Spatial Analysis: Modeling in a GIS Environment. GeoInformation International, Cambridge, UK.

Morris, R.E. and T.C. Myers (1990). User's Guide for the Urban Airshed Model. Volume I: User's Manual for UAM (CB-IV). Prepared for the U.S. EPA (EPA-450/4-90-007A) by Systems Applications, Inc., San Rafael, CA.

Peuker, T. K., R. J. Fowler, J. J. Little, D. M. Mark (1978). The Triangulated Irregular Network. In: Proceedings of the DTM Symposium, American Society of Photogrammetry-American Congress on Survey and Mapping.

Openshaw, S. (1984). The Modifiable Areal Unit Problem. CATMOG 38, Geo Books, Norwich, England.

Quattrochi, D.A. and M.F. Goodchild (1997). Scale in Remote Sensing and GIS. CRC Lewis Publishers, New York.

Rao, S.T. (1987). Application of the Urban Airshed Model to the New York Metropolitan Area. Prepared for the U.S. EPA (EPA 450 4-87-011).

Scheffe, R.D. (1990). Urban Airshed Model Study of Five Cities (Summary Report). Prepared for the U.S. EPA (EPA-450/4-90-006A).

Shimp, D.R. and S.G. Campbell (1996). Using a Geographic Information System to Evaluate PM10 Area Source Emissions. California Air Resources Board, Sacramento, CA.

Sierra Research, Inc (1993). SJVAQS/AUSPEX Agriculture Emissions Inventory. Prepared for San Joaquin Valleywide Air Pollution Study Agency. Report No. SR93-04-01.

Skieist, I. (1977) (eds). Handbook of Adhesives. Van Nostrand Reinhold Company, New York.

U.C. Cooperative Extension (1996). Sample Production Cost Estimates by Crops. Department of Agriculture Economics, Cooperative Extension, University of California, Davis, CA 95616. Or contact the county Farm Advisor.